Summer 2008

Interactions of the Fragile X Protein Family with the G Quadruplex Forming RNA Targets

Lakshmi Menon

INTERACTIONS OF THE FRAGILE X PROTEIN FAMILY WITH THE G QUADRUPLEX FORMING RNA TARGETS

A Dissertation

Submitted to the Bayer School of

Natural and Environmental Sciences

Duquesne University

In partial fulfillment of the requirements for

the degree of Doctor of Philosophy

By

Lakshmi Menon

June 2008

INTERACTIONS OF THE FRAGILE X PROTEIN FAMILY WITH THE G QUADRUPLEX FORMING RNA TARGETS

By

Lakshmi Menon

Approved June 26, 2008

_____ Mihaela Rita Mihailescu, Ph.D. Assistant Professor of Chemistry and Biochemistry (Dissertation Director)	_____ Charles T. Dameron, Ph.D. Associate Professor of Chemistry and Biochemistry Saint Francis University (External Reviewer)
_____ Ellen Gawalt, Ph.D. Assistant Professor of Chemistry and Biochemistry (Committee Member)	_____ Nancy Trun, Ph.D. Associate Professor of Biology (External Reviewer)
_____ David W. Seybert, Ph.D. Dean, Bayer School of Natural and Environmental Sciences Professor of Chemistry and Biochemistry (Committee Member)	_____ Jeffry D. Madura, Ph. D. Chair, Department of Chemistry and Biochemistry Professor of Chemistry and Biochemistry

ABSTRACT

INTERACTIONS OF THE FRAGILE X PROTEIN FAMILY WITH THE G QUADRUPLEX FORMING RNA TARGETS

By

Lakshmi Menon

June 2008

Dissertation was supervised by Professor Dr. Mihaela Rita Mihailescu

The absence of fragile X mental retardation protein (FMRP) causes Fragile X Syndrome, the most common form of inherited mental retardation. FMRP, an RNA binding protein and, part of large ribonucleoprotein complexes, is proposed to participate in the transport, localization and translation regulation of specific target mRNAs. How the absence of this protein leads to mental retardation is not clearly understood. FMRP has been shown to use its Arginine-Glycine-Glycine (RGG) box to bind with high affinity to target RNA sequences containing G quadruplex structures. Two of the proposed mRNA targets of FMRP are the human semaphorin 3F (S3F) mRNA and the microtubule associated protein 1B (MAP1B) mRNA. The overall objective of this project is to understand the role played by the G quadruplex structure in the FMRP recognition

of its mRNA targets. In this study, we demonstrate that S3F/MAP1B RNAs, fold into parallel intramolecular G quadruplex structures, to which the FMRP RGG box domain binds with high affinity and specificity. We also analyzed the interactions between the S3F/MAP1B RNA targets and the two autosomal paralogs of FMRP, the fragile X related protein 1 and 2.

DEDICATION

To my mom and grand mom - for providing me with their constant source of motivation and their enormous emphasis on the value of education. To Sandeep, my husband - without you, pursuing a doctoral degree would just have been a dream. To my immediate family, who all along have rallied behind and encouraged me every step of the way.

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to my thesis advisor Dr. Rita Mihailescu for her continuous encouragement, optimism and confidence in me. Her confidence in me regularly challenged me to perform my research to the fullest extent of my capabilities. I am grateful for the countless number of hours she spent proofreading my research papers, discussing inexhaustible number of ideas to investigate, and improving my dissertation. Her office door was always open to our research problems and she never once hesitated to write me a recommendation letter for the innumerable number of competitions that peaked my interest. Apart from her technical guidance, Dr. Rita has been kind enough to provide me with her personal guidance and insights. I will always remember her patience, support and kindness. Thank you Dr. Rita.

When I go to conferences and people ask me questions, I realize now that I am at ease in answering them. This would not have happened, if it had not been for my committee members. I would like to thank every one of my committee members, Drs. David W. Seybert, Nancy Trun, Charles T. Dameron and Ellen Gawalt for their generous time and the promptness in their feedback. My committee members set a high standard for my research early on, and showed me how to maintain them. Every one of them helped me through the different phases of this learning process, patiently watching over my progress. Their everlasting encouragement and genuine care has taught me to be always persistent and keep achieving higher goals. My appreciation for their constant guidance is immeasurable.

A very special thank you note to Dr. Evanseck, who guided me through a tough time, and did not doubt my abilities.

To Dr. Madura - His continuous dedication to his students is both admirable and inspirational. I feel honored to have had a chance learn from him.

I consider myself very lucky in having had an opportunity to work with Medhavi Bole and Joshua Lipay. My deepest gratitude especially to Medhavi for the incredible amount of help she has given me over the last three years. A special thanks to Dr. Kristina O. Pazehoski, Tyler C. Collins, Rebekah Dedrick, Dawn Bisi, Carolynne Ricardo, Marielle Balili and Dr. Yasser Shaban for being such good friends. Their good humor and insights helped me keep going. I would also like to extend my thanks to my fellow lab members for their incredible support - Kimberly Zanotti, Patrick Lackey, Julie Wong Chong, Timothy L. Evans and Kailey Ruszkowski.

To Ian Welsh, Sujata Patil, Daniel Bodnar, David Hardesty, Lance Crosby, Mary Jo Babinsack, Amy Stroyne and Sandy Russell for their tremendous patience and assistance.

Thank you to the Department of Chemistry and Biochemistry, Duquesne University and all my friends.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

2-AP: 2-aminopurine

CGG : Cytosine guanine guanine

D_2O: deuterium oxide

DLS : dimer linkage sequence

DLS_9AP: 2-AP incorporated in the 9^{th} position of the dimer linkage sequence RNA

EDTA: ethylenediamine tetraacetic acid

EMSA: electrophoretic mobility shift assays

E.coli : Escherichia coli

FXS : fragile X syndrome

FMR1: fragile X mental retardation 1 gene

FMRP: fragile X mental retardation protein

FRDA : friedreich ataxia

FXR1P: fragile X related protein 1

FXR2P: fragile X related protein 2

HCV : hepatitis C Virus

K_d: equilibrium dissociation constant

KH: heterogeneous nuclear ribonucleoprotein K Homology

MAP1B RNA: microtubule associated protein 1B RNA

MAP1B-19AP RNA: microtubule associated protein 1B RNA in which the 2-AP reporter is located at position 19

RGG box: arginine-glycine-glycine rich domain

RG cluster: arginine-glycine cluster

SELEX : systematic evolution of ligands by exponential enrichment

S3F RNA: human semaphorin 3F RNA

S3F-M2 RNA: Mutated semaphorin 3F RNA

S3F-M2_15AP RNA: mutated semaphorin 3F RNA with 2AP reporter located at position 15.

S3F-sh: S3F RNA without the stem structure

NES: nuclear export signal

NLS: nuclear localization signal

NDF: N-terminal domain of FMRP

TBE: Tris, Boric acid, EDTA pH 8.3

UTR: untranslated region

$\Delta G°$: standard Gibbs free energy change

$\Delta H°$: standard state enthalpy change

$\Delta S°$: standard state entropy change

$\Delta C^0_{p\,obs}$: change in molar heat capacity

T_H : temperature at which the enthalpy does not make any contributions to the energy of the system

T_S : temperature at which the entropy do not make any contributions to the energy of the system

Chapter 1

1. Introduction

1.1 Fragile X Syndrome - A new form of mutation

An association between sex linked inheritance and mental retardation had long been known, but the first description of sex-linked familial mental retardation came in 1943, when Martin and Bell described a family with sex-linked mental retardation without dysmorphic features (1). In 1969, Lubs discovered that there was an unusual secondary constriction at the end of the long arm of the X-chromosome in affected males and in their unaffected female relatives (2). These results were impossible to reproduce until the importance of the folate-deficient, thymidine-deficient medium, which was used in the initial studies to culture lymphocytes, was realized (3). Since the 1960s and early 1970s, progress toward mapping the gene involved has been steady and rewarding, and in 1991 the precise genetic defect that causes fragile X syndrome (FXS) was characterized (4). In the principles of genetic inheritance, it was initially thought that mutated genes would be transmitted intact to the next generation, however, when the gene responsible for FXS was cloned in 1991, a new concept of gene mutation emerged, where an unstable repeat expansion serially mutates upon transmission (5).

Fragile X Mental Retardation Syndrome (FXMR/ FXS) is the most common form of inherited mental retardation next to Down's syndrome (6-8). This syndrome was named after the cytogenetically visible fragile site present on the X-chromosome at q27.3 site (9). With a prevalence of one in 4000 males and one in 8000 females, FXS has a large impact on the world population (10). It has also been reported that ~30% of individuals with the fragile X syndrome have autism (8). Mild to moderate cognitive and behavioral abnormalities accompanied by macroorchidism (enlargement of testis) and subtle craniofacial dysmorphia are some of the clinical manifestations of this syndrome (4, 10, 11). Individuals with both FXS and autism have lowered cognitive abilities, language problems, and behavioral difficulties compared to those with only FXS.

At the molecular level, FXS is caused by the transcriptional silencing of the *fmr1* gene (12). The resulting effects range from learning difficulty and hyperactivity to severe mental retardation in individuals.

1.2 Fragile X Mental Retardation 1 gene (*fmr1*) family

The progressive unstable expansion of $(CGG)_n$ repeats and the subsequent hypermethylation of the CpG island in the 5'-untranslated region (5'-UTR) of the *fmr1* gene causes the transcriptional inactivation (4,13) of this gene and eventually the absence of FMR1 protein (which has been named the fragile X mental retardation protein- FMRP) (13,14), which is now accepted as the basis for the phenotype. The complexity of this syndrome is increased by the fact that the CGG repeats are not pure, having an AGG interruption every 9-10 repeat units, the most common allele found being $(CGG)_{9\text{-}10}$ AGG $(CGG)_9$ AGG $(CGG)_9$. The loss or the lack of this interruption has been proposed to

correlate with the CGG repeat instability (5). FMRP is absent in the majority of fragile X patients (14). The length of the trinucleotide expansion correlates with the clinical severity of the effect (15). In normal individuals, this triplet repeat is polymorphic in length and content, with 7-52 triplets and a mean of 30. The premutated individuals have between 60-200 repeats, with an abnormal level of *fmr1* mRNA being found in these individuals. Neurological symptoms may develop in males and occasionally females with permutated alleles who, with aging, could develop tremor, ataxia, peripheral neuropathy, and cognitive deficits. Atrophy of the brain and the white-matter disease is usually seen in these cases (16).

Finally, in individuals with full mutational phenotypes, a massive expansion of the Cytosine Guanine Guanine (CGG) repeat occurs, beyond 230 triplets and usually exceeding 700. Triplet repeat amplification also occurs in fragile XE MR (FRAXE), friedreich ataxia (FRDA), myotonic dystrophy (MD)**, s**pinocerebellar ataxia type 8 (SCA8)**,** spinocerebellar ataxia type 12 (SCA12), spinobulbar muscular atrophy (also called the kennedy disease) and huntington disease. The amplification in each of these diseases can occur in both sexes at each generation. Like Fragile X, there are a threshold number of triplet copies required for the symptoms to be visible (17).

As a consequence of the triplet expansion in FXS, the upstream CpG island is concomitantly methylated. This abnormal methylation indirectly attracts the enzyme histone deacetylase, the chromatin conformation of the *fmr1* gene is altered and the result is the transcriptional silencing of the gene (4,10,11).

The *fmr1* gene spans about 38 kilobases (kb) and encodes a 4.4 kb transcript consisting of 17 exons (18, 19). This gene is highly conserved in vertebrates and belongs to a family that also contains two autosomal paralogs named *fxr1* and *fxr2* (fragile X related genes). There is only one homolog for the three genes in *Drosophila melanogaster,* called the *dfxr* (20).

The alternative splicing of *fmr1* gene could result in 20 possible FMRP isoforms but only five or six protein isoforms with molecular weights ranging from 70 to 80 kDa have been identified, *in vivo* (21). The major form has a molecular weight of 78 kDa (22).

1.3 The Fragile X Mental Retardation Protein (FMRP) localization

The *fmr 1* gene encodes for the RNA-binding protein, fragile X mental retardation protein in normal individuals (23). FMRP expression is widespread in human and murine tissues (24). *In situ* hybridization experiments with adult mouse tissues revealed that this protein is found in abundance in the neurons, particularly those of the hippocampus and Purkinje cells of the cerebellum (22) and in the testes, ovary, esophageal epithelium, thymus, eye and spleen with a moderate expression in colon, uterus, thyroid and liver. FMRP is not expressed in the heart, aorta or muscle (11).

In the cell, FMRP is predominantly found in the cytoplasm in association with poly (A) mRNP complexes derived from the polyribosomes (11, 25). The nuclear localization of the isoforms lacking the sequence encoded by exon 14 and the identification of the conserved nuclear localization and export signals (NLS and NES) in its sequence suggests that FMRP is a putative nucleocytoplasmic shuttling protein (26).

1.4 Possible cellular functions of FMRP

The function of FMRP has been extensively studied in the past several years, however, the cellular role of this protein and how its loss causes mental retardation are poorly understood. This remains one of the major issues in understanding the biology of FXS. FMRP is proposed to regulate mRNA transport and translation in a manner critical for the development of the neurons; hence, the loss of this protein would cause defects in the proper translation of many mRNAs, explaining the pleiotropic phenotype associated with FXS (10,11). It is believed that FMRP is involved in a number of interactions with cytoplasmic nucleic acids, protein partners and ribonucleoprotein (mRNP) particles in actively translating ribosomes, to achieve this function (figure 1.1 A). Although studies have shown that FMRP can regulate translation negatively both *in vitro* and *in vivo*, the exact role played by this protein remains elusive (11, 27-31). More recently, it has been proposed that in fact FMRP may undergo a change of function, from translation repressor to activator, in response to a neurotransmitter stimulation event (32, 33).It has also been suggested that FMRP has nucleic acid chaperone properties (34). The significant conclusion of all these studies was that this protein seems to exert its function through RNA binding; thus, it is important to define the RNA binding properties of FMRP and identify its mRNA targets.

1.5 FMRP is an RNA binding protein

FMRP binds to RNA homopolymers *in vitro* with a preference for poly (G) > Poly (U) > poly (A)> poly(C) (35). Independent sequence analysis of this 632 amino acid protein by Siomi *et al*. (36) and Ashley *et al*. (19), revealed the presence of two types of RNA binding motifs namely the K homology domains (KH) and one arginine-glycine-glycine rich region (RGG box) (4,10,11,26). In addition, a NDF motif (N-terminal domain of FMRP) has been isolated in the N-terminus of FMRP (26).

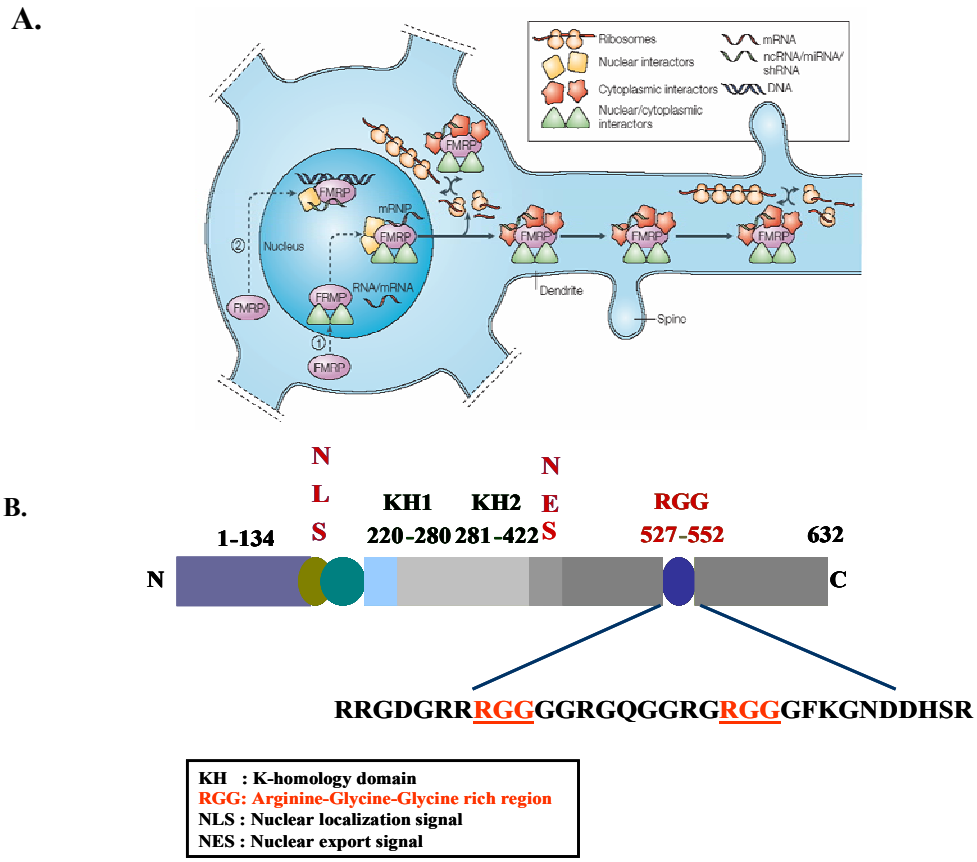

Figure 1.1(A) Proposed model for neuronal functions of FMRP, where it enters the nucleus, interacts with other proteins and mRNA targets to form the ribonucleoprotein complexes. FMRP then is involved in the export of mRNA from the nucleus to the cytoplasm (36). **(B)** Schematic representation of the fragile X mental retardation protein (FMRP) domains.

The KH domain is highly conserved, being found among other RNA-binding proteins such as hnRNP K, yeast MER-1 splicing regulator, Sam 68 and chicken vigillin (37,38). The RGG box was first isolated in hnRNP U protein. This RNA binding domain was proposed to be a predictor of RNA binding activity (39).The RGG box has been shown to autonomously bind to homopolymeric RNA in hnRNP U, hnRNP A1, nucleolin, and fibrillarian (38,39). The RGG Box has been thought to bind to RNA non-specifically.

Initial *in vitro* homopolymer binding assays performed with translated FMRP and carboxyl terminal truncated FMRP proteins suggested that the first KH domain is involved in RNA recognition, whereas the RGG box played a role in increasing the protein binding affinity for RNA (26,37). It has been reported that the FMRP also interacts with small noncoding RNAs such as BC1 and BC200, which in turn mediates its binding to the specific mRNAs (40). This theory is controversial, with recent studies in five laboratories indicating that the interactions between the BC1 RNA and the FMRP could not be reproduced either in the *in vitro* or the *in viv*o conditions (41, 42). The progress in the understanding of the RNA-binding properties of FMRP was furthered by the identification of mRNA targets (10,11).

1.6 mRNA targets of FMRP contain guanine rich sequences proposed to fold into G quadruplex structures

Although reports available suggest that regions of FMRP have independent roles in the RNA recognition, the RGG box seems to be pivotal for FMRP: RNA interaction. The interactions of FMRP with the mRNA targets have been mapped to the RGG box,

which has been shown to be sufficient for binding *in vitro* (43-47) and necessary for the assembly of the RNP complex. Schaeffer and colleagues provided the first evidence that purified FMRP RGG box has a very high affinity for RNA sequences that have been proposed to adopt a G quadruplex structure (43). They also suggested that since this specific and high affinity binding site in the FMRP mRNA interacts with FMRP *in vitro* and *in vivo*, the FMRP may be involved in regulating its own translation (20,43,48).

Darnell and colleagues using the SELEX (systematic evolution of ligands by exponential enrichment) method, showed that sequences proposed to fold into G quadruplex structures serve as high-affinity targets for FMRP. Seventy one FMRP potential mRNA targets have been identified in this study and interestingly, the mRNAs identified bound to the RGG box of FMRP and not its KH domain . Most of the proteins encoded by these mRNAs, are important for neuronal functions (44).

In a combined study, Brown *et al*. (49) used a microarray strategy to identify *in vivo* targets of FMRP from mouse brain and also compared the mRNA polysomal profile from normal and fragile X patient lymphoblast cells. 14 mRNAs were found be common among the two pools. To identify the mRNAs present in the close vicinity of FMRP in neurons, Miyashiro *et al*. performed *in situ* cDNA synthesis and the comparison of the results from Brown *et al*. with those of Miyashiro, revealed 17 mRNAs in common (50). The most interesting finding was that the mRNAs identified had the potential to form G quadruplex structures, suggesting that these structures are physiologically relevant to FXS (10).

1.7 G quadruplexes are the foundation for the self-assembly of guanosine

For several years now, one of the questions repeatedly raised on G quadruplexes is: do these structures have a biological relevance? Is there a direct proof of their existence in living cells? There are no known examples of G quadruplexes *in vivo,* but there is growing evidence that DNA and RNA G quadruplex structures could form *in vitro* at chromosomal telomeres, gene promoter regions, recombination sites, RNA packaging sites, and RNA dimerization domains (51, 52). The RNA G quadruplexes have been proposed to have a number of individual roles including structural (53) translational regulation (54, 55), intronic splicing (54, 56) and in protein binding (46, 47, 57). The RNA G quadruplex structures have also been shown experimentally to be more stable than the DNA G quadruplexes (57). A search of the human genome suggests that about 3000 genes contain G quadruplex motifs in the 5'UTR. These regions have been proposed to play a regulatory role in controlling transcription (58, 59).

Edges that are self complementary hydrogen-bond donors and acceptors, polarizable aromatic surface and a strong molecular dipole make guanosine associate with itself and ideal for stacking (52,59). In 1962, Gellert and co-workers proposed the G tetrad structure for guanosine gels. They developed a model, from fiber diffraction studies, that showed the N1-H and N2-H donor atoms of one guanine molecule pairs with the N7 and O6 atoms of a neighboring guanine molecule to form a G quartet (60). The G quartet is defined as the planar association of four guanines in DNA (and RNA), held together by Watson-Crick base pairing and Hoogsteen hydrogen bonds (figure 1.2) (61).

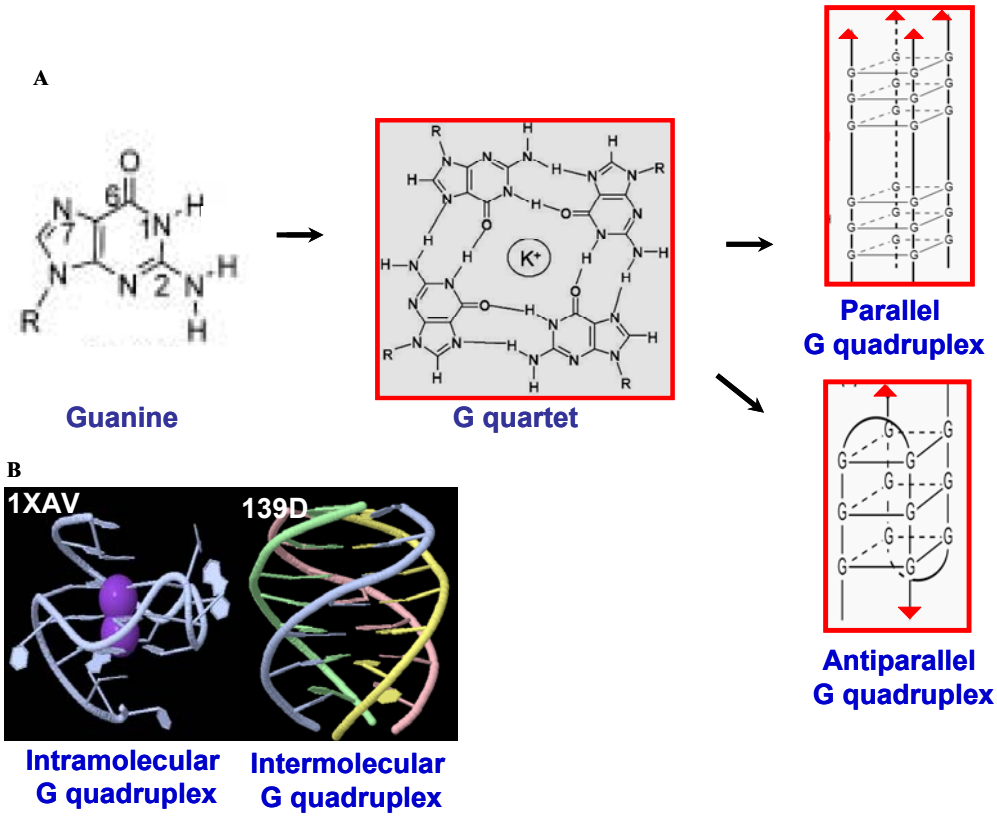

Figure 1.2 Structure of G quadruplexes. Adapted from Han and Hurley, 2000 (61). **(A)** Shows the schematic representation G quartet formed through Hoogsteen hydrogen bonding. The G quartets stack to form parallel or anti-parallel G quadruplex structures. **(B)** Intramolecular G quadruplex sequences with either four G rich repeats or long G tracts folded upon themselves. Intermolecular G quadruplexes where sequences containing two or more G rich repeats form G-G hairpins, which in turn dimerize to form several types of stable bimolecular quadruplexes. The PDB identities are shown in the figure inset.

Hydrophobic stacking of several G quartets 3.25 Å apart result in the formation of G quadruplexes (52). Nucleic acids rich in guanine sequences form inter- or intra molecular G quadruplexes. Cations like K^+ or Na^+ form cation-dipole interactions with the eight guanine residues and are located between the two quartets. The quartet stacking is stabilized by reducing the repulsion between 2 x 4 oxygen atoms and enhancing the hydrogen bond strength (63). The G quartet structures are disrupted by the presence of Li^+.

1.8 Human Semaphorin 3F and Microtubule Associated Protein 1B RNA - proposed mRNA targets of FMRP

In this study, we have chosen two proposed mRNAs rich in G-content, namely the human semaphorin 3F RNA and the microtubule associated protein 1B RNA, that have both been identified *in vivo* and *in vitro* as potential mRNA targets of FMRP. Moreover, both these RNAs have been proposed to interact with FMRP via the recognition of a G quadruplex present in their structure by the protein RGG box domain (32, 44-47).

1.9 Human Semaphorin 3F mRNA (S3F)

The human semaphorin 3F RNA encodes for Sema 3F, a putative secreted protein belonging to the semaphorin family of proteins, largely identified as signaling proteins involved in neuronal development. The members of this family have chemoattractant and repulsion functions and each has the 500 amino acid signature, the semaphorin domain (64-66). The Sema 3F protein affects axon and cell guidance in the developing nervous system and may also play a role in the synaptic reorganization in the adult brain. The expression of the *SEMA 3F* gene has also been reported to suppress tumor formation in nude mice and to cause the alteration of the cellular response to drugs inducing apoptosis. In addition, this protein has been suggested to play an important role in the immune and cardiovascular systems (64,66,67). It has been shown that FMRP binds to semaphorin mRNA sequence *in vivo* (68). FMRP has also been shown to bind human semaphorin mRNA with a high affinity (K_d = 75 nm) *in vitro* (44). In addition, this RNA shows a decrease in the altered polysome association in fragile X cells (68). The G rich sequence

of this mRNA could fold into G quadruplex structures and does not show multiple secondary structures when folded with the RNA structure 4.11 software.

1.10 Microtubule Associated Protein 1B RNA (MAP1B RNA)

The microtubule associated protein 1B (MAP1B) mRNA encodes for MAP1B, a protein that is expressed first during the brain development process and is involved in modulating microtubule dynamics, neurite extension and synapse development (69). There is abundant evidence supporting the proposal that MAP1B mRNA is a relevant *in vivo* FMRP target. FMRP has been shown to bind to the MAP1B mRNA *in vitro* with a K_d = 505 nM. The MAP1B RNA also shows increased altered polysome association in fragile X cells (44). Ramos *et al*. showed that this RNA binds to the FMRP RGG Box *in vitro* (45). The *Drosophila* homolog of FMRP reduces the levels of Futsch, a microtubule-associated protein homolog of the mammalian MAP1B (70). An *in vivo* study by Lu *et al*. showed that FMRP expression governs the translation of MAP1B RNA: when the levels of FMRP expression increased in the hippocampus, the levels of MAP1B RNA were found to be down regulated. *In vivo* studies also show that FMRP negatively regulates the translation of MAP1B and controls the microtubule stability in brain neuron development (40, 71). The lack of FMRP has been shown to misregulate the MAP1B translation affecting the dynamic organization of neuronal cytoskeleton, and leading to abnormal microtubule stability (40, 71). In addition, FMRP has been shown to co-localize with ribosomes, ribosomal RNA and MAP1B mRNA (72).

1.11 Fragile X Mental Retardation related proteins (FXR1P and FXR2P)

FMRP has two autosomal paralogs, namely the FXR1 and FXR2 proteins (FXR1P and FXR2P), the three proteins collectively forming the fragile X related protein family (73). The amino acid sequence analysis revealed that the two autosomal paralogs have ~60% amino acid identity, with regions of 90% sequence identity to FMRP (73, 74). FXR1P and FXR2P are also cytoplasmic RNA-binding proteins, each containing two KH domains. The FXR2P is divergent from FMRP and FXR1P in the C-terminal region, in that it has an RG cluster instead of an RGG box. The FXR proteins have been found to be associated predominantly with the ribosomal 60S subunit, and the finding that they have similar RNA binding domains lead to the suggestion that the FXR1P and FXR2P might compensate for the FMRP function (75). However, the comparison of the expression levels of each of these proteins in different tissues and cellular distributions suggests that each of the FXR proteins might have an independent function (76).

Two other FMRP-associated proteins namely, the nucleolin and NUFIP (nuclear FMRP-interacting protein) have also been identified (47, 77). Although the role played by these two FMRP interacting proteins or their physiological significance is not yet known, it has been proposed that they may function as part of the FMRP-RNP complex and regulate specific RNA metabolism.

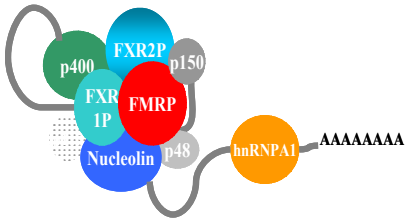

Figure 1.3 Schematic representation of FMRP interacting proteins.

1.12 Significance of this study

The important idea that has been expressed in this introduction is that the FMRP associates with polyribosomes and is involved in the translational repression of specific messenger RNA targets, important for neuronal development. The loss of this protein would cause defects in the proper translation of many mRNAs, thus explaining the pleiotropic phenotype associated with FXS. The specific mechanism by which FMRP recognizes its target RNAs and regulates their translation remains elusive, and has been a subject of continuous interest. This project aims to understand the molecular basis of FMRP recognition of its mRNA targets and the role played by the G quadruplex structure in this process. Beyond the context of the Fragile X Syndrome, this study will provide additional information about the G quadruplex forming RNA molecules and their interactions with proteins. Also, the properties of the RGG box RNA binding motif are characterized in the context of *specific* RNA binding.

Goal 1. To understand the principles of recognition between G quadruplex forming S3F/MAP1B mRNAs and the RGG Box of FMRP by using structural and thermodynamic approaches.

The mechanism by which FMRP recognizes its target RNA has been a subject of continuous interest. A model for FMRP function suggests that this protein has RNA-binding activity, associates with polyribosomes and is involved in the translational repression of specific messenger RNA target, important for neuronal development. Since the role of FMRP as a translation regulator is entirely dependent upon its ability to interact with and act upon its RNA target, one focus of this study is to understand the

principles of recognition between the RGG Box of the FMRP and its mRNA targets, the human semaphorin 3F RNA and the MAP1B RNA. We employ thermodynamic and structural methods here to determine if the stability and the secondary structure of semaphorin 3F/MAP1B RNA targets are affected by their interaction with the RGG Box of FMRP. This aim is of particular interest in the context of fragile X research, because this would contribute to our ability to identify other *in vivo* targets of the FMRP.

Goal 2. To define the role played by RGG box in protein-RNA interactions by comparing the binding of FMRP RGG box to S3F/MAP1B mRNA with that of the RGG boxes from other RNA-binding proteins.

Recently, FMRP was found to bind with intramolecular G quadruplex forming RNA via its RGG box (32, 44-47). The RGG boxes were initially thought to increase the RNA-binding affinity of proteins by non-specific interactions. However, Darnell *et al*. reported that the binding between the FMRP RGG box and G quadruplex forming RNA is specific though the role of RGG box in these interactions remains unclear (44). In this goal, we investigate if the recognition of the G quadruplex structures is a unique feature of the FMRP RGG Box. The role played by RGG box specific G quadruplex recognition was evaluated by comparing the binding activity of the FMRP RGG Box with the RGG box of the Fragile X retardation protein 1 (FXR1P). The interaction of the RG cluster of FXR2P with the Human S3F and MAP1B RNA targets was also analyzed.

Chapter 2

2. Materials & Methods

2.1 Expression and Purification of the T7 RNA polymerase

The expression vector pT7-911Q containing the Histidine-tagged T7 RNA polymerase gene (a gift from Dr. John Marino, NIST) was first transformed in *Escherichia coli* BL21 (DE3) competent cells. These cells were plated on an LB (Luria Bertani) agar plate containing 100 μg/mL ampicillin and were grown overnight at 37°C. A single colony was picked from the plate and inoculated into 5 mL of LB containing 5 μg/mL ampicillin. The starter culture was grown at 37°C with shaking at 250 rpm. A total of 500 μL of an overnight culture of *E. coli* BL21 (DE3) cells were diluted into 500 mL of LB medium containing ampicillin (50 μg/mL). The culture was grown at 37 °C until optical density at 600 nm reached 0.4-0.6. The culture was then induced with 500 μL of 1M IPTG (isopropyl-β-D-thiogalactopyranoside) and the incubation was continued for an additional 3-4 hours in the same temperature until the optical density of the cell culture reached 0.8. The cells were harvested by centrifugation at 6000 *g* for 10 minutes at 4°C using an Avanti ™ J-20 XP high-performance centrifuge (Beckman Coulter™). The pelleted cells were then stored in a 50 mL tube and placed in the -80°C freezer until further use. The wet weight of the cells was determined and on average the yield was 6mg /liter of cell culture. The cells were frozen-thawed four times and then resuspended

in 25 mL of lysis buffer (50 mM Tris pH 8.0, 0.1 M NaCl, 5.0 mM β-mercaptoethanol and 5% glycerol) containing 1 mM imidazole. 2 mg of lysozyme along with 150 µL of DNAse in 5 mL of 10X buffer containing 100mM Tris, 25 mM magnesium chloride and 5 mM calcium chloride, pH 7.5 were added to digest the cellular membranes and the released DNA. The final volume was adjusted to 50 mL with lysis buffer and the reaction mixture was placed in the -20°C freezer overnight. The following day, the suspension was centrifuged at 18,000 *g* for 30 minutes at 4°C to separate the cell debris.

The soluble fraction containing recombinant His-tag T7 RNA polymerase was loaded onto a Ni-NTA column (Qiagen, Germany) and pre-equilibrated with lysis buffer containing 1 mM imidazole). Unwanted proteins were washed out with the lysis buffer containing 10 mM, 50 mM imidazole. The T7 RNA polymerase was eluted using lysis buffer containing 100 mM imidazole. The protein purity was then confirmed by the presence of a single band with a molecular weight of ~100 kDa on a 15% SDS-PAGE gel. For protein visualization, the gels were stained with Coomassie brilliant blue and destained with destaining buffer (40% ethanol or methanol, 10% acetic acid and water made up to 100 mL). The purified protein was mixed with glycerol (25%) and sodium chloride (300 mM) and then stored at −80 °C, frozen in liquid nitrogen.

2.2 *In vitro* transcription reactions of the unlabeled RNA samples

The unlabeled RNA oligonucleotides (S3F-lg, S3F-M2, MAP1B and Munc-13 site 1) were synthesized by *in vitro* transcription reactions using T7 RNA polymerase (produced in-house), following the procedure by Milligan & Uhlenbeck (78). The synthetic DNA templates were purchased from Trilink Biotechnologies, Inc.

Several 40 µL transcription reactions were performed to optimize conditions prior to the large scale transcription reaction. Typically 10 individual reactions were performed where only one condition (such as the concentration of T7 RNA polymerase, magnesium chloride or the concentrations of each nucleotide) was varied. Basic transcription reaction conditions for the large scale reaction were: 40 mM Tris (pH 8.3), 5 mM dithiothreitol, 1 mM spermidine, 0.01% (v/v) Triton X-100, 25 mM magnesium chloride. The concentrations for the nucleotides were varied according to their occurrence in the RNA sequence. In the case of MAP1B RNA we used 8 mM GTP, 2 mM ATP, 2 mM UTP and 4 mM CTP; for the semaphorin 3F RNA the concentrations of the nucleotides were 7 mM GTP, 3 mM ATP, 3 mM UTP and 3 mM CTP. The reaction mixtures were adjusted to a final volume of 10 mL with distilled water. The DNA templates were prepared by mixing 1.1 mM C-TOP (5'CTA ATA CGA CTC ACT ATA G 3') and 1 mM DNA for the respective RNA sequences, S3F-lg 5'-GGG AGG GGG GUG AUU GGA AGG GAG GGA GGU GGC C UUC C-3'; S3F-M2 RNA 5'-GGC UGG UGA UUG GAA GGG AGG GAG GUG GCC AGC C -3'; MAP1B RNA 5'- GGC GCU GGG AGA GGG CGG AGG GGG AGG CGG CGC C-3' and boiled for 5 minutes and cooled in the same water bath until the temperature reached 37°C to allow the formation of primer-template junction. The reaction mixture was incubated with 0.1 mg/mL T7 RNA polymerase for 6-8 hours at 37°C. The reaction was then arrested by the addition of 1.6 mL EDTA from a 500 mM stock at pH 8.0 followed by 500 µl from a stock of 3 M sodium acetate and three and a half volumes (~ 45 mL) of ethanol. The reaction mixture was then stored at -20°C for 24 hours.

The RNA oligonucleotides were purified using 20% denaturing gel (acrylamide/bisacrylamide/8 M urea) at 1200 V, 200 mA, 90 W. The RNA bands, which were visualized by UV shadowing at 254 nm, were cut and the RNA was extracted by electrophoretic elution, followed by extensive dialysis against 10 mM Tris (pH 7.5) or 10 mM cacodylic acid (pH 6.5). All experiments were performed with RNA samples that were annealed by heating at 95°C, followed by slow cooling at room temperature for 20 minutes.

2.3 Peptide Synthesis

The FMRP RGG Box (528-RRGDGRRRGGGGRGQGGRGRGGGFKGNDDHSR-560), FXR1P RGG Box (433-HQRDSRRRPGGRGRSVSGGRGRGGPRGKSS-463), the FXR2P RG cluster (475-TRGEESRRRPTGGRGRGPPPAPRP-500) and the hepatitis C virus core peptide (PRRGPRLGVRATRKTSERSQPRGRRQPIPKVRHQTGRRGSRPNWGPNDPRRRSRNLGK) were chemically synthesized and purified by the Peptide Synthesis Unit at the University of Pittsburgh, Center for Biotechnology & Bioengineering.

2.4 UV Spectroscopy

The UV melting curves of the unlabeled S3F-lg, S3F-M2, MAP1B and of the 2-AP labeled S3F-M2_15AP, MAP1B-19AP RNAs were measured using a Varian Cary 3E spectrophotometer equipped with a Peltier cell. The RNA samples were annealed in the standard buffer 10 mM cacodylic acid, pH 6.5 or 10 mM Tris, pH 7.5, containing either 150 mM KCl or 150 mM LiCl in a final volume of 200 μL. The experiments were carried

out in 10 mm path-length quartz cuvettes and the RNA samples were heated from 20°C to 99°C at a rate of 0.2°C/minute, recording points every 1°C. To prevent evaporation, mineral oil (Sigma-Aldrich) was layered on top of the RNA samples. Blank samples were treated in the same manner. Depending upon the RNA concentration, the spectral absorbance was measured either at 295 or 305 nm, wavelengths that have been previously identified to be sensitive to G quadruplex dissociation (79). The OD ranges of the RNA samples at either wavelengths were between 0.1 – 0.8.

To determine if S3F-M2 RNA or MAP1B RNA fold into an *intermolecular* or *intramolecular* conformation, we measured their melting temperatures at different RNA concentrations in the range 10 - 80 μM for S3F-M2 RNA and 3 – 100 μM for MAP1B RNA. For *intermolecular* species with *n* number of strands, $1/T_m$ depends linearly on the natural logarithm of the total RNA concentration (c_T):

$$\frac{1}{T_m} = \frac{R(n-1)}{\Delta H^0_{vH}} \ln c_T + \frac{\Delta S^0_{vH} - (n-1)R\ln 2 + R\ln n}{\Delta H^0_{vH}} \quad (1)$$

where R is the gas constant and ΔH^0_{vH} and ΔS^0_{vH} are the van't Hoff thermodynamic parameters.

For *intramolecular* species, T_m is independent of the total RNA concentration c_T:

$$\text{n=1 and } \frac{1}{T_m} = \frac{\Delta S^0_{vH}}{\Delta H^0_{vH}} \quad (2)$$

To determine the thermodynamic parameters of G quadruplex formation, the transition of the G quadruplex dissociation in S3F-M2, S3F-M2_15AP, MAP1B and MAP1B-19AP RNA was fitted assuming an independent two state model:

$$A(T) = \frac{A_U + A_F \; e^{\frac{-\Delta H^0}{RT}} \; e^{\frac{\Delta S^0}{R}}}{e^{\frac{-\Delta H^0}{RT}} \; e^{\frac{\Delta S^0}{R}} + 1} \quad (3)$$

where A_U and A_F represent the absorbance of the unfolded and native G quadruplex RNA, respectively, and R is the universal gas constant.

The number of potassium ion equivalents released upon the melting of the MAP1B G quadruplex structure was determined by measuring the melting curves of this RNA in the presence of increasing concentrations of KCl, in the range 0.1 mM – 10 mM.

The RNA: RGG peptide complexes were formed by mixing the RGG boxes of the FMRP, FXR1P, or the RG cluster of FXR2P with the S3F-M2 RNA or the MAP1B RNA in a 1:1 ratio, and allowed to equilibrate for 30 minutes. The melting temperature of the complex of S3F-M2 and the MAP1B with various RGG boxes was determined using the same parameters described above for the free RNA. All melting curves of the S3F-M2: RGG peptide and the MAP1B RNA: RGG peptide complexes were corrected for the peptide contribution by subtracting the melting curves of the corresponding free peptides.

2.5 Circular Dichroism (CD) Spectroscopy

The CD spectra were recorded on a Jasco J-810 spectropolarimeter at 25°C. A 1 mm path-length quartz cuvette (Starna Cells) was used for all experiments. The G quadruplex formation in the S3F-M2 RNA or the MAP1B RNA structure (at a 10 μM concentration in 10 mM cacodylic acid, pH 6.5) was monitored by titrating increasing amounts of KCl from a 4 M stock solution to a final concentration of 150 mM for S3F-M2 RNA and 10 mM KCl for MAP1B RNA. The spectra were measured between 200-350 nm and corrected for solvent contributions and dilutions. Each spectrum was scanned

3 times with a 1 s response time and a 2 nm bandwidth. For the binding studies, increasing amounts of the RGG peptides of FMRP, FXR1P, or the FXR2P RG cluster (0-100 μM) were titrated into a fixed concentration of RNA (10 μM) in 10 mM cacodylic acid, pH 6.5, and 150 mM KCl for S3F-M2 RNA and 10 mM cacodylic acid, pH 6.5, and 10 mM KCl for MAP1B RNA. The CD spectra were recorded after each addition of the peptide and the interactions were monitored by measuring the molar ellipticity of the RNA at 264 nm. All spectra were corrected by subtracting the free RGG peptide contributions at each peptide concentration.

To determine if the addition of a 1:10 RNA: RGG peptide ratio results in the degradation of the RNA, proteinase K (1μg) (which degrades the RGG peptides) was added to the 1:10 RNA: RGG peptide complex for 1 hour at 25°C and the CD spectra were re-recorded and corrected for the presence of proteinase K.

2.6 Fluorescence Spectroscopy

The S3F-M2_15AP RNA used in the fluorescence spectroscopy studies was constructed by replacing the adenine at the 15th position in S3F-M2 with the fluorescent analog 2-aminopurine (2-AP) (Dharmacon, Inc). MAP1B-19AP RNA was constructed by incorporating 2-AP (Dharmacon, Inc) at the 19th position. The pure 2-AP labeled RNAs were resuspended in sterile deionized water.

Steady state fluorescence spectroscopy measurements of S3F-M2_15AP and MAP1B-19AP RNA were performed on a J.Y. Horiba Fluoromax-3 with a 150 W ozone-free Xenon arc lamp, equipped with variable temperature control in the sample chamber. Experiments were carried out in 3 mm path-length quartz cuvettes (Starna Cells), on

samples with a final volume of 150 µl. The excitation wavelength was at 310 nm and the emission spectrum was recorded in the range of 330 - 450 nm and the bandpasses for the excitation and emission monochromators were set at 5 nm.

The binding of the FMRP RGG box to the 2-AP labeled RNAs was measured by titrating increasing concentrations of the peptide (15 nM from a 10 µM stock solution in the case of S3F-M2_15AP and 60 nM from a 10 µM stock solution for MAP1B-RNA) to a fixed concentration of 150 nM S3F-M2_15AP or 400 nM of MAP1B RNA. These samples were equilibrated for 15 minutes after each peptide addition. The volume changes in the sample due to each peptide addition varied from 0.6% for the first point to 6% for the last point. Duplicates or triplicates were performed for each experiment and the standard errors reported for the data uncertainty from the best-fit theoretical curves. These errors are usually larger than the standard uncertainty of the measurements. The same procedure was repeated for the FXR1P and the FXR2 RG cluster. The binding dissociation constant, K_d, was determined by plotting the normalized 2-AP steady-state fluorescence intensity measured at 371 nm as a function of the FMRP RGG box concentration and then fitting the binding curves to the equation:

$$F = 1 + \left(\frac{I_B}{I_F} - 1\right) \bullet \frac{(K_d + [P]_t + [RNA]_t) - \sqrt{(K_d + [P]_t + [RNA]_t)^2 - 4 \bullet [RNA]_t \bullet [P]_t}}{2 \bullet [RNA]_t} \qquad (4)$$

where I_F and I_B represent the steady-state fluorescence intensities of the free and bound S3F-M2_15AP or MAP1B-19AP RNA, $[RNA]_t$ is the total concentration of S3F-M2_15AP or the MAP1B-19AP RNA, and $[P]_t$ is the total RGG box or RG cluster peptide concentration.

Competition experiments were performed by monitoring the binding of the FMRP, FXR1P or RGG peptides to S3F-M2_15AP in the presence of a 10-fold excess of unlabeled G quadruplex forming Munc-13 site 1 RNA (5'-GGGUUUGAGGCCAUGGG AGGCAAGGAGCUGGAC-3') or in the presence of a 6-fold excess of the FXR2P RG cluster peptide.

Competition experiments for the binding of FMRP RGG Box to MAP1B-19AP RNA were performed in the presence of a 10-fold excess of unlabeled non-specific G quadruplex forming Munc-13 site 1 RNA, or in the presence of non-specific single-stranded rComp4 RNA (5'-AUUAUUAUUAUUA-3') or DLS_9AP (5'-UCACGGCGAPGCUGUGA-3'), a generous gift from Dr. Bruce Armitage (Carnegie Mellon University). We also performed competition experiments in the presence of a 10-fold excess of the non-specific Hepatitis C virus peptide (HCV).

The thermodynamic parameters for the FMRP RGG box binding to S3F-M2_15AP and MAP1B-19AP RNA were determined by measuring the $K_{obs}=1/K_d$ at different temperatures in the range 20 - 45°C. When the standard enthalpy, ΔH^0_{obs}, and entropy, ΔS^0_{obs}, do not depend on temperature, the van't Hoff plot is linear and shows dependence of $\ln(K_{obs})$ upon 1/T. In this case, the standard enthalpy and entropy of binding were then determined from the slope and intercept of the graph:

$$R \ln K_{obs} = \Delta S_b^0 - \frac{1}{T} \Delta H_b^0 \qquad (5)$$

The van't Hoff plot for the FMRP RGG box binding to the MAP1B RNA is non-linear, since ΔH^0_{obs} and ΔS^0_{obs} depend on temperature. The entropy and enthalpy changes at different temperatures were calculated by obtaining the heat capacity change of the

system, $\Delta C^0_{p\,obs}$ and the temperatures, T_H and T_S, (the temperatures where $\Delta H^0_{obs} = 0$ and $\Delta S^0_{obs} = 0$, respectively), at which the enthalpy and entropy do not make any contributions to the Gibbs free energy of the system:

$$\Delta H^0_{obs} = \Delta C^0_{p\,obs}(T - T_H) \tag{6}$$

$$\Delta S^0_{obs} = \Delta C^0_{p\,obs} \ln(T / T_S) \tag{7}$$

The heat capacity change $\Delta C^0_{p\,obs}$ and the T_H and T_S values for the non linear van't Hoff plot were determined from the fit of the following equation (80, 81):

$$\ln K_{obs} = \frac{\Delta C^0_{p\,obs}}{R}\left[\frac{T_H}{T} - \ln\frac{T_S}{T} - 1\right] \tag{8}$$

The temperature dependence of the change in the standard enthalpy and entropy (ΔH^0_{obs} and ΔS^0_{obs}) at various temperatures, was determined using equations 6 and 7.

2.7 NMR Spectroscopy

The one dimensional (1D) ^{1}H NMR spectra of S3F-M2 RNA , MAP1B RNA or Munc 13 site 1 RNA were acquired at 29°C on a 500 MHz Varian Unity Plus spectrometer or on a 500 MHz Bruker AVANCE^{TM} spectrometer. The water suppression was accomplished using the jump-and-return pulse sequence (82) with the maximum of excitation set at 11 ppm or using the Watergate pulse sequence (83). S3F-M2 RNA (387 μM), MAP1B RNA (500 μM) or Munc 13 site 1 RNA (263 μM) were prepared in 10 mM Tris (pH 7.5) at a 90% H_2O/10%D_2O ratio. To observe the G quadruplex formation, each sample was titrated, with increasing concentrations of KCl up to 150 mM from a 4 M stock solution for S3F-M2 RNA, and up to 10 mM KCl for MAP1B RNA.The melting of the S3F-M2 RNA and the MAP1B RNA stem structures was monitored by recording the

1D ^{1}H NMR spectrum at different temperatures in the range of 20-60°C for S3F-M2 and 20-75°C for MAP1B RNA. These experiments were performed on a Bruker AVANCE^{TM} 500 MHz NMR spectrometer.

For the binding studies, the FMRP RGG box was titrated into the S3F-M2 RNA or the MAP1B RNA in increasing ratios up to 1:10, allowing the sample to equilibrate for 20 minutes after each peptide addition.

2.8 Electrophoretic Mobility Gel Shift Assay (EMSA)

EMSA reactions were performed in a total volume of 15 µl. The RNA: peptide complexes were prepared by mixing the RGG peptides with S3F-M2 or S3F-M2_15AP / MAP1B or MAP1B-19AP RNA (10 µM) in 1:1 or 1:2 ratios. The RNA: peptide complexes were prepared by pre-incubating the RNA with the RGG peptides or the RG cluster for 20 minutes at room temperature. The free and the bound RNA were suspended in 10 mM cacodylic acid and 150 mM KCl for S3F-M2 RNA and 10 mM KCl for MAP1B RNA. The complex formation was then resolved on 15% non-denaturing acrylamide gels that were run in the presence of 75 mM KCl, at 35 V for S3F-M2 RNA and in the presence of 10 mM KCl at 35 V and 4°C for MAP1B RNA. The electrophoretic mobilities of the free RNA and the RNA: peptide complexes were visualized by UV-shadowing at 254 nm, using an AlphaImager HP (AlphaInnotech, Inc.).

Chapter 3

3.Human Semaphorin 3F mRNA

3.1 Semaphorin 3F mRNA adopts a G quadruplex structure

Human semaphorin 3F mRNA was identified to be a potential *in vivo* target of the FMRP, based on the fact that its mRNA contains a guanine rich sequence that could possibly fold into a G quadruplex structure (44). The interactions of FMRP with this G rich sequence of semaphorin mRNA have also been visualized in living mammalian cells (68). Moreover, the mutation of the GG doublets proposed to be involved in the G quadruplex formation has also been shown to abolish the RNA-protein interactions (68), supporting the idea that the FMRP recognizes the S3F RNA via a G quadruplex motif.

To demonstrate that the S3F RNA folds in to a G quadruplex structure, we first expressed and purified a 38-nt RNA, named the S3F-lg which contained the 34-nt G-rich fragment of human semaphorin 3F mRNA proposed to interact with FMRP. Four extra nucleotides (GGGA) were added at its 5'-end for transcription purposes (figure 3.1) (44).

It is well known that the G-rich nucleic acid sequences fold in to G quadruplexes in the presence of cations like K^+, by forming cation-dipole interactions with the guanine residues (52). The CD spectrum of S3F-lg folded in the presence of K^+ ions shows a positive peak at 264 nm and a negative one at 243 nm , both signatures of a type I parallel type G quadruplex (figure 3.2), confirming the presence of these structural elements in

this RNA. However, the type of fold can be confirmed only by determining the high-resolution structure of S3F-lg RNA.

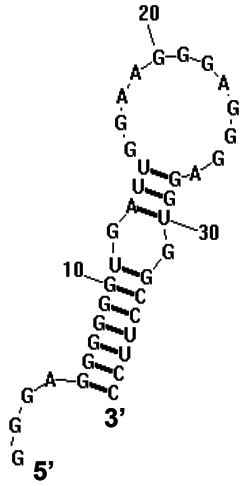

Figure 3.1 Secondary structure of the human semaphorin 3F mRNA (S3F-lg) fragment used in this study generated using the RNA structure 4.1 software (84).

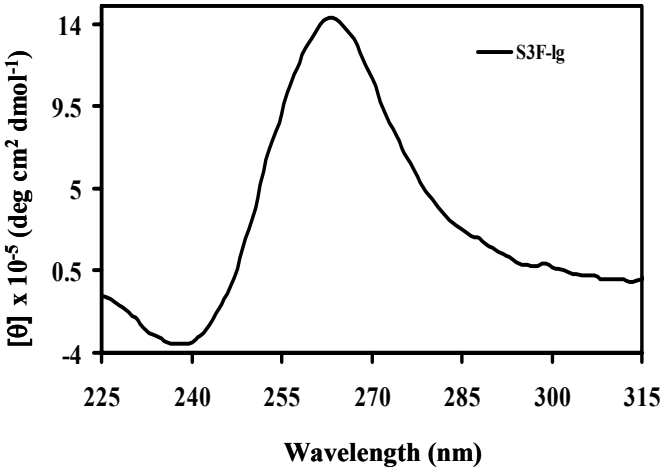

Figure 3.2 CD spectrum of 10 μM S3F-lg RNA in 10 mM cacodylic acid buffer containing 150 mM KCl, pH 6.5.

Next, we measured the UV spectroscopy thermal denaturation profile of S3F-lg RNA at 305 nm (79). The S3F-lg melting curve shows a characteristic hypochromic transition between 52°-72°C, corresponding to G quadruplex dissociation (79) (figure 3.3), indicating the presence of a G quadruplex structure in this RNA.

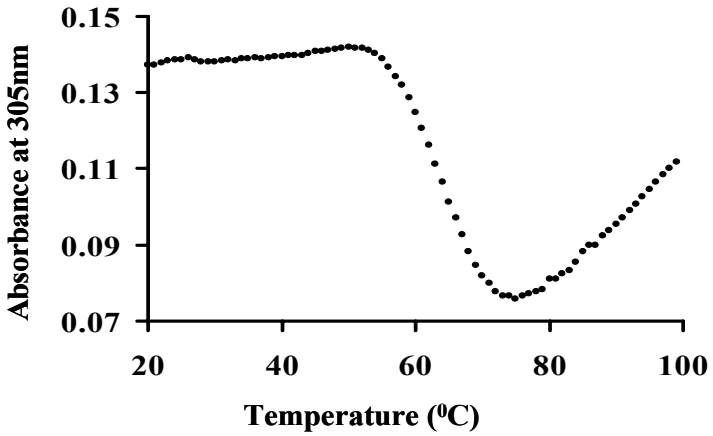

Figure 3.3 UV thermal denaturation profile of S3F-lg RNA (10μM) in 10 mM cacodylic acid buffer pH 6.5 containing 150 mM KCl.

To obtain higher resolution information about the structure of S3F-lg RNA, we used one-dimensional (1D) ^{1}H NMR spectroscopy. Very broad resonances corresponding to imino protons involved in G quartets were observed in the 10-12 ppm proton region (figure 3.4). The broadening of these resonances is due to exchange of the RNA between different conformations. No imino proton resonances were observed in the 12-14 ppm region that corresponds to Watson-Crick base pairs, indicating that the stem structure proposed in figure 3.1 does not exist in S3F-lg (figure 3.4). The addition of the extra four nucleotides (GGGA) contributes to uninterrupted stretch of ten purines at the beginning

of the S3F-lg sequence, possibly promoting the folding of this RNA into an alternate G quadruplex structure.

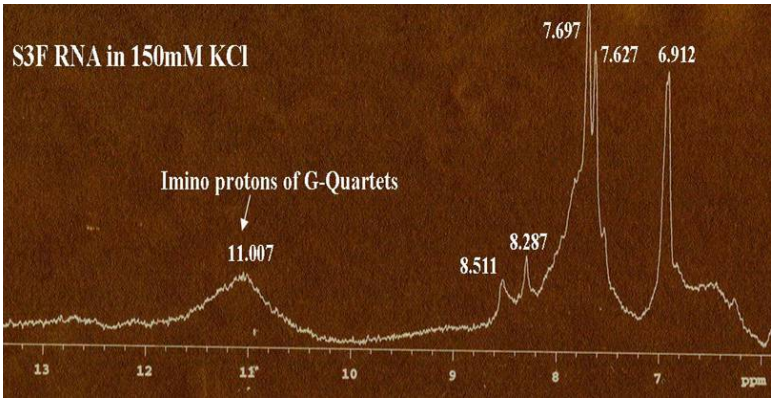

Figure 3.4 1D ^{1}H NMR spectrum of S3F-lg RNA (263µM) in 10 mM Tris pH 7.5 and 150 mM KCl concentration.

To determine if S3F-lg RNA forms alternate G quadruplex structures *in vitro* we used native gel electrophoresis, and indeed two conformations were observed on a 15% native gel performed in the presence of 75 mM KCl at all RNA concentrations investigated (figure 3.5).

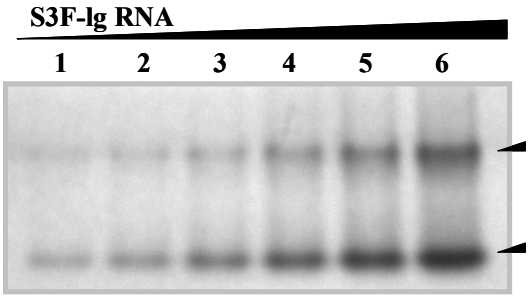

Figure 3.5 15% native gel electrophoresis performed in the presence of 75 mM KCl at different S3F-lg RNA concentrations: 1.5 µM (lane 1), 6 µM (lane 2), 10 µM (lane 3), 20 µM (lane 4), 30 µM (lane 5), 50 µM (lane 6).

To promote the folding of S3F-lg RNA into a single conformer we first removed the four GGGA nucleotides added only for transcription purposes (44). Next, we introduced specific point mutations at positions 3 (G to C) and 4 (G to U) and at the complementary positions 31 (U to A) and 32 (U to C), respectively (labeled in blue in figure 3.6) to interrupt the stretch of six consecutive guanines in the region proposed to fold into a stem. This mutated RNA was named S3F-M2 RNA.

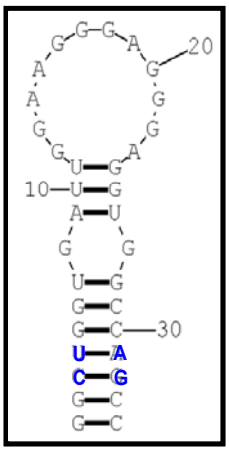

Figure 3.6 Secondary structure of the S3F-M2 RNA fragment with the mutated base pairs highlighted in blue.

To determine if S3F-M2 RNA maintains the ability to form a G quadruplex structure, we first used CD spectroscopy. The CD spectrum of S3F-M2 RNA showed the spectral features of a type I G quadruplex CD spectrum, when increasing concentrations of KCl were titrated into a fixed concentration of S3F-M2 RNA (10 μM in 10 mM cacodylic acid buffer), namely a positive band at 263 nm and a negative band at 238 nm (figure 3.7).

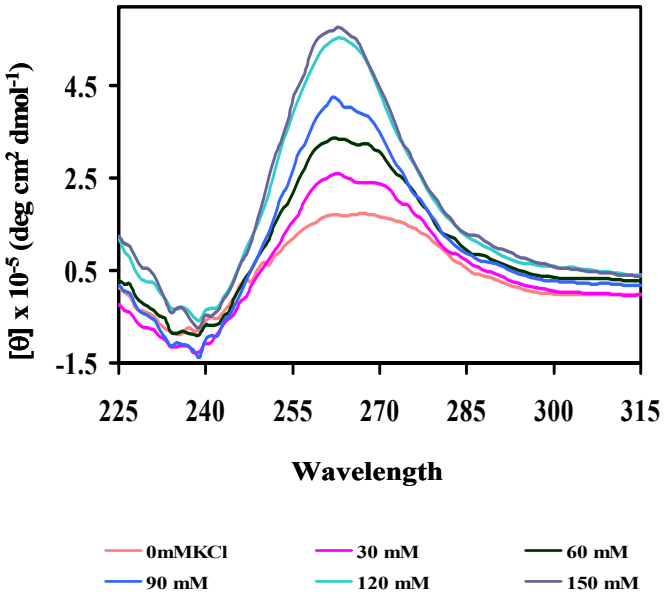

Figure 3.7 CD spectra showing the G quadruplex formation in S3F-M2 RNA by titrating increasing concentrations of KCl.

The native gel electrophoresis of S3F-M2 RNA indicated that at concentrations lower than 10 µM this RNA exists in a single conformation, whereas at higher concentrations S3F-M2 adopts more conformations.

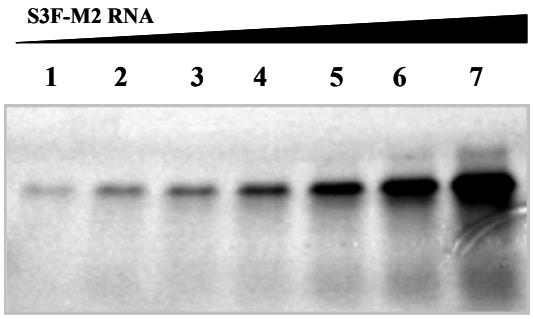

Figure 3.8 Native gel of different concentrations of S3F-M2 RNA. Lane 1 - S3F M2 RNA (3µM), Lane 2 - 6µM, Lane 3 - 10µM,Lane 4 - 20µM Lane 5 - 30µM, Lane 6 - 50µM,Lane 7 - 100µM

We also monitored the formation of the S3F-M2 G quadruplex by 1D ^{1}H NMR spectroscopy. Resonances corresponding to imino protons involved in G quartets, as well as resonances corresponding to Watson-Crick base pairs, indicated the presence of both a stem and a G quadruplex in the structure of this RNA (figure 3.9). The imino proton resonances of the G quadruplex were very broad, consistent with an exchange between the different S3F-M2 conformers formed at the high RNA concentration required when using this technique. Resonances of the Watson-Crick imino protons also become much broader upon addition of increasing KCl concentrations supporting the evidence that the RNA is exchanged between different conformations. These findings did not allow us to pursue high resolution NMR spectroscopy studies of the S3F-M2 RNA structure.

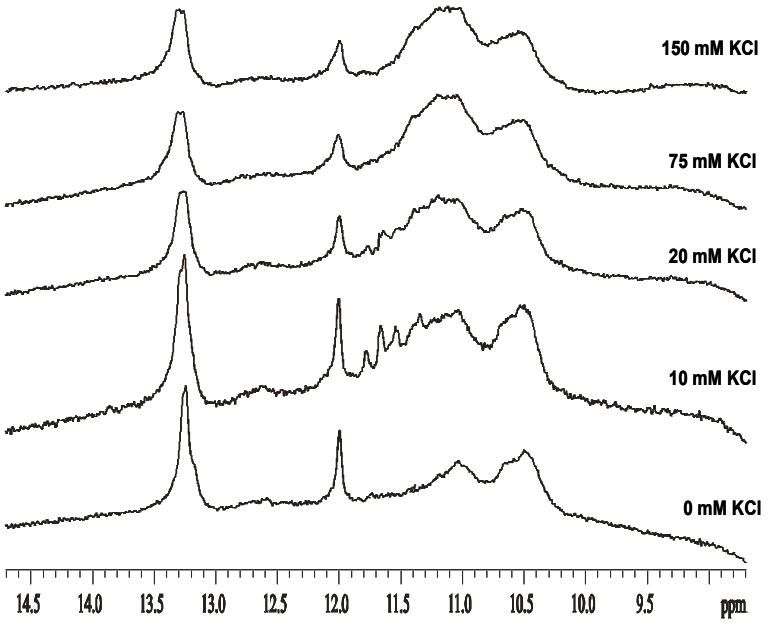

Figure 3.9 1D ^{1}H NMR spectrum of S3F-M2 RNA (387μM) in 10 mM Tris pH 7.5 and various KCl concentrations.

3.2 Thermodynamics of G quadruplex formation in S3F-M2 RNA

To obtain the thermodynamic parameters of G quadruplex formation in S3F-M2 RNA we employed UV spectroscopy. The UV thermal melting profile of 10 µM S3F-M2 RNA folded in the presence of 150 mM KCl shows a hypochromic transition between 38-67ºC (indicated in red in figure 3.10A), and a hyperchromic transition starting around 75°C. We assigned the 38-67°C hypochromic transition with a melting point ~ 52°C, to the S3F-M2 RNA G quadruplex dissociation (79).This transition was not observed when the RNA was folded in the presence of 150 mM LiCl (figure 3.10B), since G quadruplexes do not form in the presence of Li^+ ions.

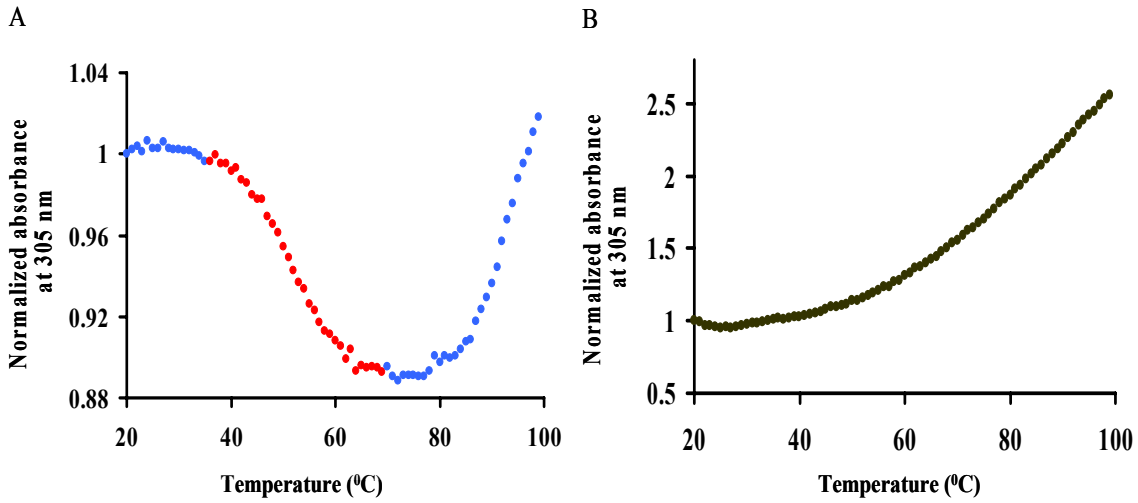

Figure 3.10 UV thermal denaturation profile of S3F-M2 RNA (10µM in 10 mM cacodylic acid buffer, containing either 150 mM KCl **(A)** or 150 mM LiCl **(B).**

First, we measured the reversibility of the melting curves of S3F-M2 RNA in the range 25-70°C and 25-99°C, respectively (figure 3.11 A and B) to rule out the possibility that the hyperchromic transition starting around 75°C is due to the RNA degradation at high temperatures. As seen in figure 3.11 these curves are superimposable.

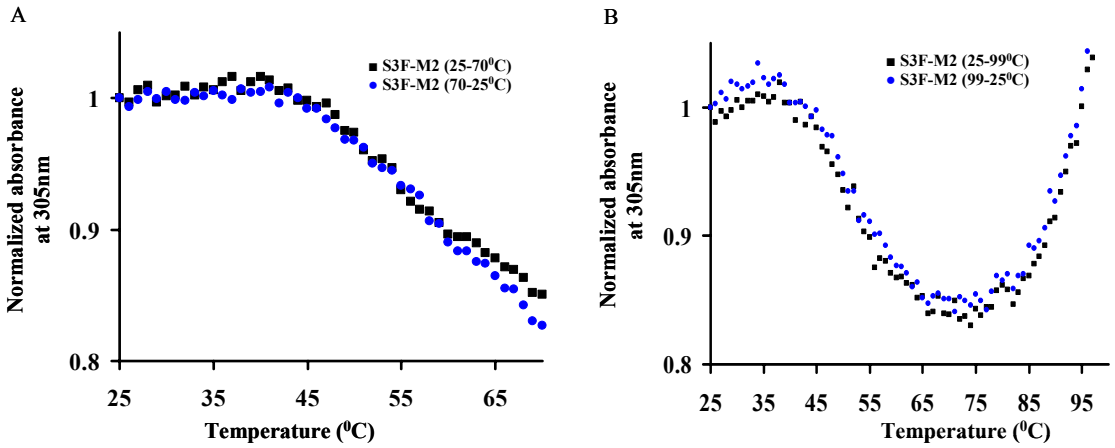

Figure 3.11 Reversible UV melting profiles of S3F-M2 RNA (10μM) containing 150 mM KCl from **(A)** 25-70°C and 70-25°C **(B)** 25-99°C and 99-25°C.

We hypothesized then that the hyperchromic transition starting at 75°C in the presence of K^+ corresponds to the melting of the S3F-M2 stem structure (47). To test this hypothesis we constructed a S3F RNA from which the stem region has been removed (S3F-sh). First, we established that S3F-sh RNA maintains the ability to form a G quadruplex structure, as evidenced by its type I CD spectrum (positive band ~265 nm and negative band ~ 240 nm) and the presence of G quartet imino proton resonances in its 1D ^{1}H NMR spectrum (data not shown). The UV thermal denaturation profile of S3F-sh RNA measured at 305 nm shows a 40-65°C hypochromic transition, corresponding to a G quadruplex melting point of ~52°C (figure 3.12). However, the hyperchromic transition starting around 65°C is still present in the UV melting profile of S3F-sh, which

lacks a stem structure (figure 3.12) suggesting that the hyperchromic transition observed in the UV melting profile of S3F-M2 RNA, is not due to the melting of its stem structure. The S3F-M2, S3F-sh and S3F-lg RNAs have an uninterrupted stretch of fourteen purines (starting at G12 for S3F-M2 RNA- figure 3.6). It is possible that upon the melting of the G quadruplex structure, the liberated rG residues stack with their rA nearest neighbors, and these rG-rA stacks will melt at increasing temperatures, giving rise to the hyperchromic transition observed above 65°C (46, 84,85).

The comparison of the melting points of the G quadruplex structures formed by the S3F-sh, S3F-M2 and S3F-lg RNAs shows that the S3F-sh forms a similar G quadruplex structure (T_m~ 52°C) to S3F-M2 RNA. This result indicates that the mutations introduced in the stem of S3F-M2 RNA do not affect the ability of the G quadruplex forming sequence to fold into a structure similar to that of the wild type S3F-lg.

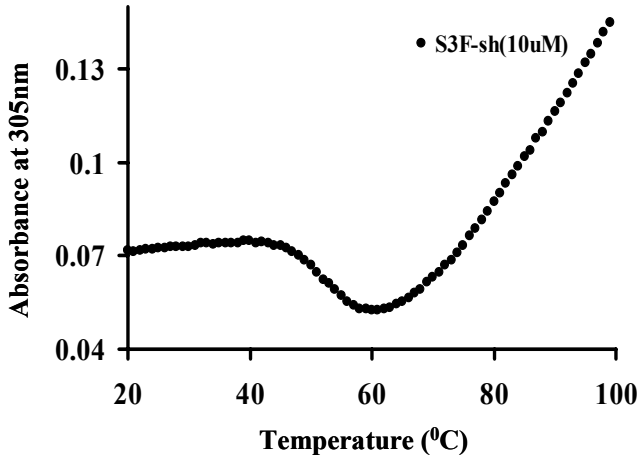

Figure 3.12 UV thermal denaturation profile of S3F-sh RNA (10μM) in 10 mM cacodylic acid buffer, pH 6.5 containing 150 mM KCl.

To determine if S3F-M2 RNA forms an *intramolecular* or an *intermolecular* G quadruplex structure, we measured its melting temperature at various RNA concentrations in the range 10-80 µM. At concentrations of RNA lower than < 10 µM, S3F-M2 RNA favors the formation of a single species, with a melting temperature T_m of ~52°C (figure 3.13 A blue trace) whereas at higher RNA concentrations, a second hypochromic transition appears in the range 63-86°C, corresponding to a new S3F-M2 conformation with a melting temperature of ~79°C (figure 3.13 A red trace). The T_m of the 38-67°C hypochromic transition is independent of the RNA concentration (figure 3.13 B), indicating that the G quadruplex conformation formed by S3F-M2 at low RNA concentrations is *intramolecular*.

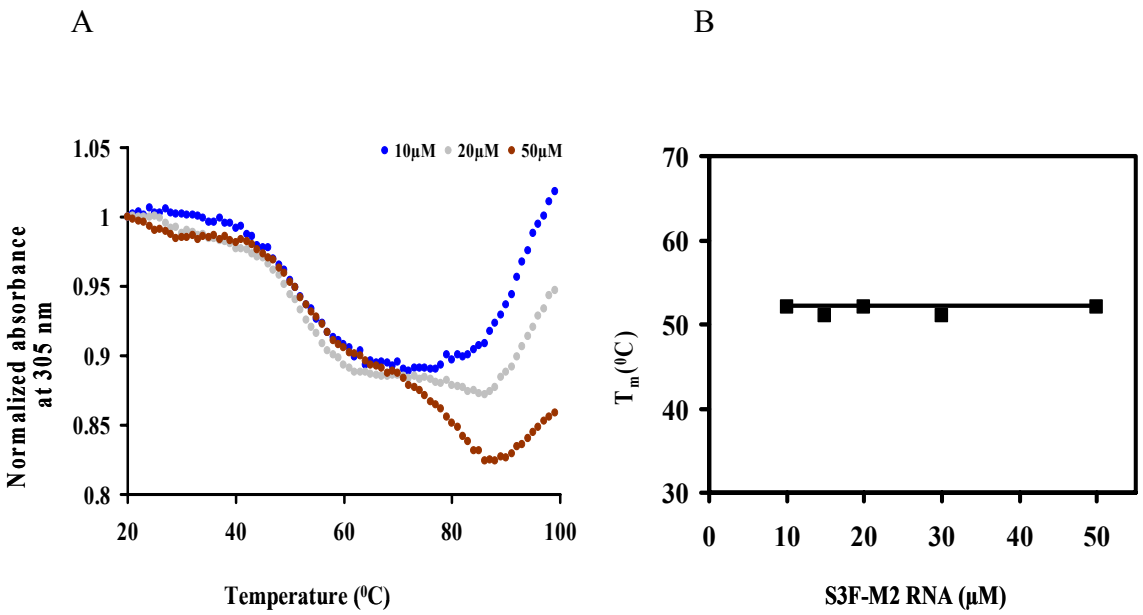

Figure 3.13 (A) UV-melting profiles of S3F-M2 RNA in 10 mM cacodylic acid, pH 6.5 containing 150 mM KCl at the following RNA concentrations: 10 µM (blue trace), 20 µM (gray trace), 50 µM (red trace). **(B)** Plot of the melting temperature of the S3F-M2 RNA G quartet as a function of the RNA concentration.

The standard enthalpy, entropy and free energy of G quadruplex formation in S3F-M2 RNA, were obtained by fitting the 38-67°C hypochromic transition to equation 3 (materials and methods), and are summarized in table 3.1. The values of the thermodynamic parameters for G quadruplex formation (ΔH^0 = -43.1 ± 0.1 kcal/mol and ΔG^0 = -3.6 ± 0.1 kcal/mol) are consistent with the presence of two G quartet planes in the structure of S3F-M2 RNA (the enthalpy of formation of a single G quartet plane in an intramolecular G quadruplex, measured in similar experimental conditions, ranges from -18 to -25 kcal/mol (87).

Molecule	T_m(°C)	$\Delta H°_{VH}$ (kcal/mol)	$\Delta S°_{VH}$ ($calmol^{-1}K^{-1}$)	$\Delta G°_{VH}$ at 25°C (kcal/mol)
S3F-M2	51.6 ± 0 .1	-43.1 ± 0.1	-132.1 ± 0.4	-3.6 ± 0.1

Table 3.1 Thermodynamic parameters for the formation of the G quartet structure of S3F-M2 RNA.

We have proposed a possible model of S3F-M2 RNA structure consistent with these results, in which two G-tetrads are stacked in a parallel manner (figure 3.14).

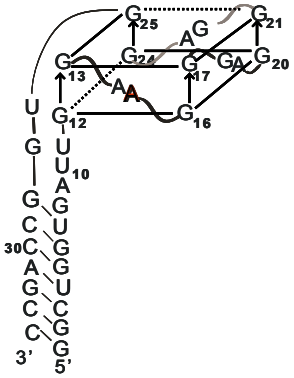

Figure 3.14 Proposed G quadruplex parallel structure of S3F-M2 RNA.

3.3 Thermodynamics of FMRP RGG box binding to S3F-M2 RNA.

To analyze the interactions of the FMRP RGG box with the G quadruplex forming S3F-M2 RNA we employed fluorescence spectroscopy. The adenine nucleotide at the 15^{th} position in S3F-M2 RNA was replaced by the highly fluorescent purine analog 2-AP (highlighted in red in figure 3.14), this RNA being named S3F-M2_15AP. Based on previous studies in our laboratory, we anticipated that the 2-AP at the 15^{th} position will be a sensitive reporter of the G quartet formation (46).

The steady-state fluorescence of 2-AP is affected by stacking interactions, and we expected to observe a change when S3F-M2_15AP RNA is folded in the presence of K^+ versus Li^+, since the structures formed by the RNA in the presence of these ions are very different. As seen in figure 3.15B the steady state fluorescence of S3F-M2_15AP increases 5-fold when the RNA is folded in the presence of KCl (forming a G quadruplex structure in which the 2-AP reporter is located in a G quadruplex surrounding loop) as compared to the case when it is folded in the presence of LiCl (that does not promote G quadruplex formation). This result establishes that the 2AP in S3F-M2_15AP RNA reports on the G quadruplex structure formation. In a control experiment we determined by electromobility shift assay that the FMRP RGG box binds identically to S3F-M2 and S3F-M2_15AP RNAs (figure 3.15A).

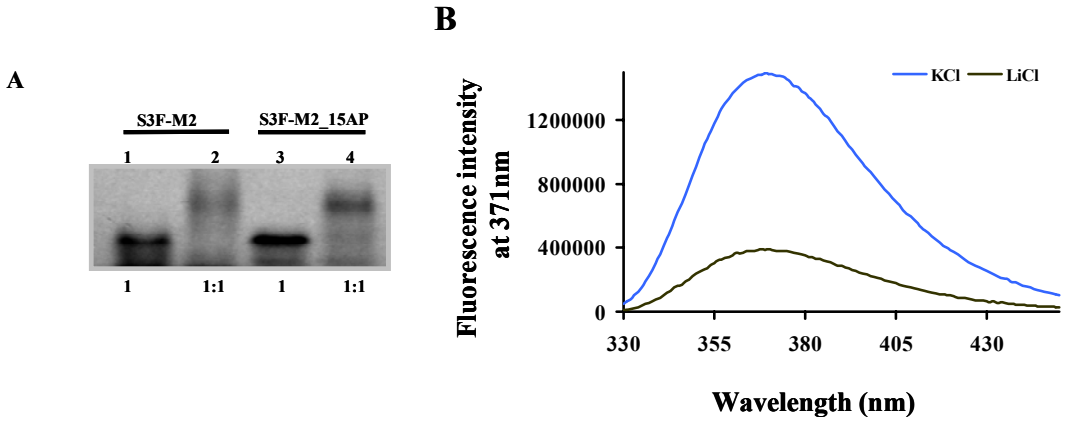

Figure 3.15 **(A)** EMSA of the FMRP RGG box binding to S3F-M2 (lanes 1 and 2) and S3F-M2_15AP RNA (lanes 3 and 4). The RNA concentration was 10 μM and the FMRP RGG box was used in a 1: 1 ratio. **(B)** The 2AP at the 15th position of the S3F-M2_15AP RNA is sensitive to the G quartet structure formation as indicated by the steady state fluorescence of S3F-M2_15AP in 150mM KCl (blue trace) or LiCl (black trace).

To obtain quantitative information about the binding of the FMRP RGG box to S3F-M2_15AP RNA, we titrated increasing concentrations of the FMRP RGG peptide to a fixed concentration of S3F-M2_15AP RNA and monitored the steady state fluorescence change of the 2-AP reporter (figure 3.16). The FMRP RGG box binding curve was fitted to equation 4 (materials and methods), determining a dissociation constant, K_d, of 0.7 ± 0.3 nM, which corresponds to a free energy of binding $\Delta G°_b$ of (-12.5 ± 0.2) kcal/mol. Our value for K_d is smaller by two orders of magnitude than the value of 75 nM reported by Darnell *et al*. (44).Two S3F conformers exist even at low RNA concentrations (figure 3.5), so it is quite possible that Darnell *et al*. measured the binding of the FMRP to both S3F conformers. In our study, we have measured the FMRP RGG box binding to the single conformer adopted by the S3F-M2 RNA at nanomolar concentrations.

Two negative control experiments were also performed. We used an RNA named Sc1-sh, for which previous reports indicate that it does not bind to the FMRP RGG box (46). Sc1-sh RNA forms a G quadruplex, but lacks a stem structure. In our experiment we titrated increasing amounts of the FMRP RGG box into a solution of 2-AP labeled Sc1-sh RNA in which the 2-AP reporter is located in one of its G quadruplex surrounding loops. In the second negative control experiment, increasing amounts of the FXR2P RG cluster, a non-binding peptide, were titrated into S3F-M2_15AP RNA.

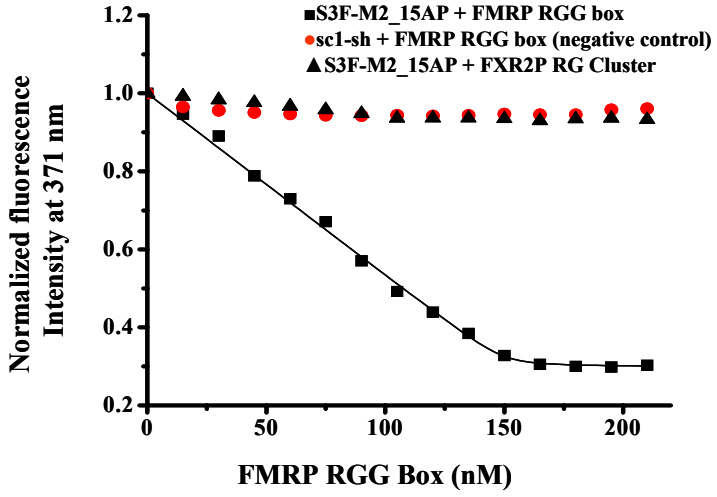

Figure 3.16 Increasing concentrations of FMRP RGG box were titrated into a solution of 150 nM S3F-M2_15AP in 10mM cacodylic acid, pH 6.5 containing 150 mM KCl. Sc1-sh RNA, which forms a G quartet but does not have a stem in its structure, has been used as a negative control (45). The binding of FMRP RGG Box to the S3F-M2 in the presence of a 6-fold excess FXR2 RG peptide is also shown.

We noted that the FMRP RGG box binds tighter to S3F-M2 RNA than to Sc1 RNA, a model system G quadruplex target of FMRP (44,46) by approximately one order of magnitude: K_d of 0.7 nM for S3F-M2 RNA versus 7 nM for Sc1 RNA. The sequence analysis of these two RNA molecules reveals differences in the G quadruplex surrounding loops and in the junction connecting the G quadruplex structure with the stem, which likely account for the difference in their binding by the FMRP RGG box.

The K_d values of the S3F-M2_15AP RNA: FMRP RGG complex were also measured in the presence of a 10-fold excess of non specific RNAs, as well as in the presence of a 6-fold excess of a non-specific peptide to determine, if the FMRP RGG box binds specifically to S3F-M2_15AP RNA. We demonstrate that the presence of a 6-fold excess of the FXR2 RG peptide does not affect the binding of the FMRP RGG box to S3F-M2_15AP RNA as seen from the value of the dissociation constant (K_d = 0.6 ± 0.3 nM) (figure 3.17). The dissociation constant values of the S3F-M2_15AP RNA-FMRP RGG complex were also measured in the presence of a 10-fold excess of Munc 13 site 1 RNA (K_d = 0.8 ± 0.5 nM) and found to be similar to those measured in the absence of the non-specific RNAs. These results indicated that the FMRP RGG box binds specifically and with high affinity to the S3F-M2_15AP RNA G quadruplex structure.

To characterize the thermodynamics of binding between the FMRP RGG box and S3F-M2_15AP RNA we have determined the enthalpy and entropy of binding by measuring the equilibrium association constant, $K_{obs}=1/K_d$, as a function of temperature. The thermodynamic parameters of binding (summarized in table 3.2) were determined from the slope and intercept of the linear Van't Hoff plot of ln (K_{obs}) versus 1/T. The

change in enthalpy (ΔH^0) in this plot is independent of temperature (equation 5 and figure 3.18).

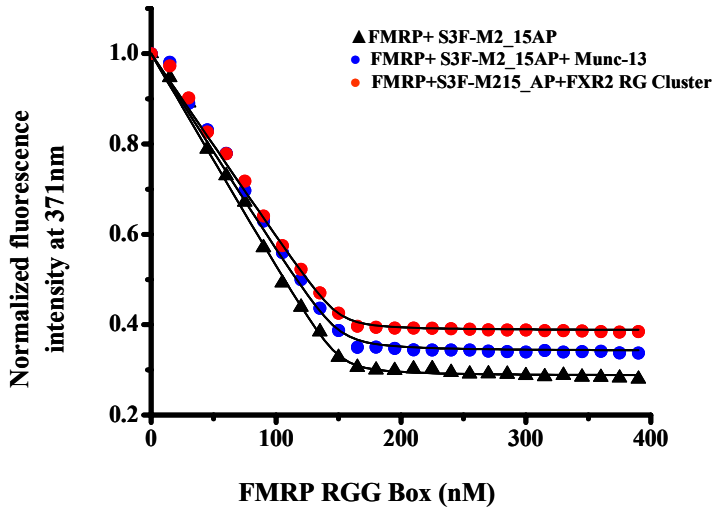

Figure 3.17 Binding of the FMRP RGG box to S3FM2_15AP in the absence (black triangles, K_d= 0.7 ± 0.3 nM) and presence (blue circles, K_d= 0.8 ± 0.5 nM) of a 10-fold excess of Munc-13 site 1 RNA or of a 6-fold excess of FXR2 RG cluster (red circles, K_d= 0.6 ± 0.3 nM).

The association of FMRP RGG box with S3F-M2_15AP RNA was found to be enthalpically driven, ΔH_b^0 = -41.4 ± 3.9 kcal/mol, with an unfavorable entropic contribution $T\Delta S_b^0$ = -28.9 ± 3.8 kcal/mol. Favorable negative enthalpy changes are generally associated with the contributions from hydrogen bonds, van der Waal's or electrostatic interactions, whereas a decrease in the conformational flexibility or the exposure of hydrophobic residues to the complex surface are associated with unfavorable entropy changes. The association between Sc1 RNA with the FMRP RGG box has also been reported to be enthalpically driven, with an unfavorable entropic change (45).

However, as shown in chapter 4, this result cannot be generalized since the FMRP RGG box binding to MAP1B RNA is enthalpically driven only at temperatures higher than 30°C (32).

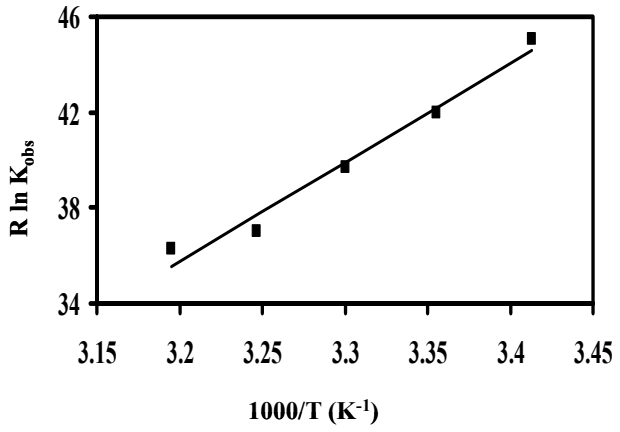

Figure 3.18 The association constant, $K_{obs}=1/K_d$ for the S3F-FMRP RGG complex was determined as a function of temperature.

Molecule	$\Delta H°_b$ (kcal/mol)	$\Delta S°_b$ ($calmol^{-1}$ K^{-1})	$\Delta G°_b$ (kcal/mol)
S3F-M2_15AP + FMRP RGG box	-41.4 ± 3.9	-96.9 ± 12.7	-12.5 ± 0.4

Table 3. 2. Thermodynamic parameters for the binding of FMRP RGG box to S3F-M2_15AP RNA.

Next, we evaluated the role played by electrostatic interactions in the S3F-M2_15AP RNA: FMRP RGG box recognition by measuring the association constant $K_{obs}= 1/K_d$ in the presence of increasing salt concentrations in the range 150-1000 mM. The dependence of K_{obs} on the concentration of monovalent salt concentrations is known as the salt dependence $\partial \log K_{obs}/\partial \log [M^+]$. We found that within experimental error there was no change in K_{obs} in the presence of increasing KCl concentrations in the range 150-1000 mM KCl (figure 3.19 and table 3.3), indicating that electrostatic contributions do not dominate the association between the FMRP RGG box and the S3F-M2_15AP RNA.

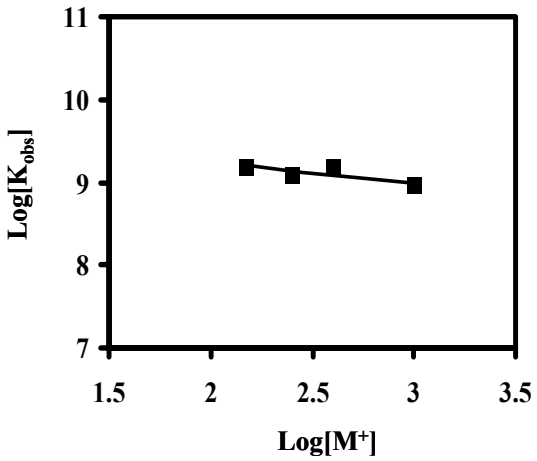

Figure 3.19 The binding of FMRP to S3F-M2_15AP RNA measured at different salt concentrations: 150 mM, 250 mM, 400 mM and 1000 mM KCl .

KCl (mM)	K_d (nM)
150	0.7 ± 0.3
250	0.8 ± 0.2
400	0.7 ± 0.2
1000	1.1 ± 0.3

Table 3.3 The dissociation constants measured for the binding of S3F-M2_15AP RNA with the FMRP RGG box at different salt concentrations.

3.4 The FMRP RGG box binding stabilizes the G quadruplex structure of S3F-M2 RNA

To determine if the FMRP RGG box binding affects the stability of the G quadruplex structure of S3F-M2 RNA, we measured its melting temperature when the RNA is in complex with the RGG peptide. The 38-67°C UV hypochromic transition corresponding to G quadruplex dissociation in the free RNA is now shifted in the range 42-78°C, corresponding to an increase of the G quartet structure T_m from ~52°C to ~65°C (table 3.4). Thus, in a 1:1 ratio, the FMRP RGG box increases the stability of S3F-M2 G quadruplex structure. The UV melting curve of the S3F-M2 RNA: FMRP RGG box complex shown in figure 3.20 has been corrected by subtracting the UV melting curve of the free FMRP RGG box peptide. Upon binding the RGG peptide, a second hypochromic transition appears in the range 79-91°C, indicating that the peptide promotes the formation of an alternate more stable structure of S3F-M2 RNA.

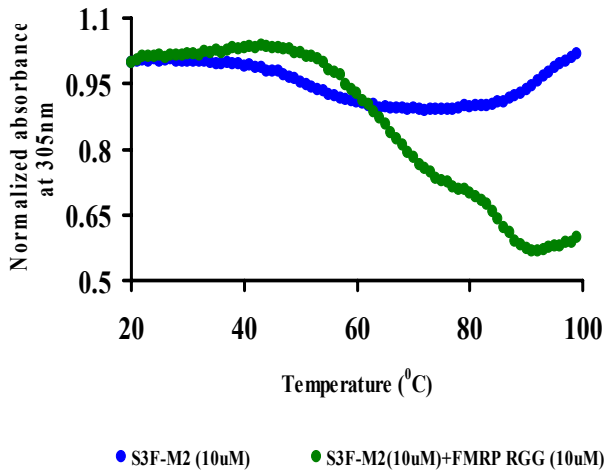

Figure 3.20 UV melting profile of 10 μM S3F-M2 free RNA (blue trace) and in a 1:1 ratio with the FMRP RGG box (green trace).

Molecule	T_m(°C)	$\Delta H°_{VH}$ (kcal/mol)	$\Delta S°_{VH}$ ($calmol^{-1}K^{-1}$)	$\Delta G°_{VH}$ at 25°C (kcal/mol)
S3F-M2	51.6 ± 0 .1	-43.1 ± 0.1	-132.1 ± 0.4	-3.6 ± 0.1
S3F-M2 + FMRP RGG	64.7 ± 0.1	-38.6 ± 0.1	-114.3 ± 0.3	-4.5 ± 0.1

Table 3.4 Thermodynamic parameters for the binding of FMRP RGG Box to the G quartet structure of S3F-M2 RNA.

The G quadruplex structure of Sc1 RNA has also been reported to be stabilized by the FMRP RGG box (46), however to a different extent. Only a small fraction of the FMRP RGG box binding free energy is used to stabilize the S3F-M2 RNA G quadruplex structure, whereas in contrast, in the case of Sc1 RNA a significant fraction of the binding free energy is used to stabilize its G quadruplex structure (47).

3.5 Interactions of the G quadruplex of S3F-M2 RNA with the Fragile X protein family

To determine if the recognition of the G quadruplex structure in the S3F-M2 RNA is a unique feature of the FMRP RGG box or if it is a property shared by the FMRP autosomal paralogs FXR1P and FXR2P, we first performed gel electrophoresis shift assays. We observed that the S3F-M2 RNA is bound by the FXR1P RGG box but not by the FXR2P RG cluster (figure 3.21, compare lanes 3 with 4 and lanes 5 with 6). The FXR1P has an RGG box domain and FXR2P has an RG cluster.

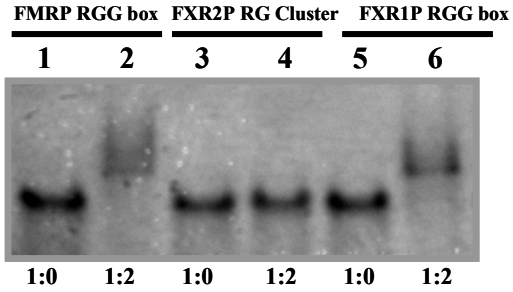

Figure 3.21 EMSA of the binding of S3F-M2 RNA by the FMRP RGG box (lanes 1 & 2), FXR2P RG cluster (lanes 3 & 4) and FXR1P RGG box (lanes 5 & 6). The RNA concentration was 20 μM and the peptides were in a 1:2 ratio with the RNA.

To quantify the binding of the FXR1P RGG box with the S3F-M2_15AP RNA we used fluorescence spectroscopy. Increasing amounts of the FXR1P RGG box were added to a fixed amount of the S3F-M2_15AP RNA (figure 3.22). The dissociation constant of the S3F-M2_15AP RNA: FXR1P RGG box complex was found to be K_d = 55.0 ± 3.8 nM, indicating that S3F-M2_15AP RNA is recognized by FXR1P with high affinity.

To determine if the FXR1P RGG box binds specifically to S3F-M2_15AP RNA we measured their binding curve in the presence of an excess of either a non-specific RNA (Munc13 RNA), or of the non-specific FXR2 RG peptide (figure 3.22). The dissociation constants in the presence of a 10-fold excess of Munc-13 site 1 RNA was found to be K_d= (55.1 ± 4.2 nM) and in the presence of a 6-fold excess of the FXR2 RG cluster was K_d= (53.8 ± 2.8 nM), indicating that the association between the S3F-M2_15AP RNA and the FXR1 RGG Box domain is specific in nature.

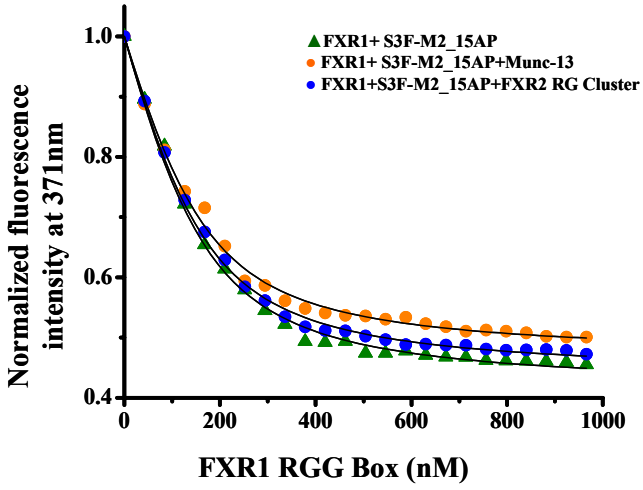

Figure 3.22 Binding of the FXR1P RGG box to S3F-M2_15AP RNA in the absence (green triangles, K_d= 55.0 ± 3.8 nM) and presence of a 10-fold excess of Munc-13 site 1 RNA (orange circles, K_d= 55.1 ± 4.2 nM) or presence of a 6-fold excess of FXR2 RG cluster (blue circles, K_d= 53.8 ± 2.8 nM).

The effects of the FXR1P RGG box and the FXR2P RG cluster on the intramolecular G quadruplex structure of S3F-M2 RNA were investigated by comparing the UV spectra of the free RNA with those of the RNA in complex with the RGG peptides. The peptides were added to the RNA in a 1:1 ratio and the melting curves were measured at 295 nm by UV spectroscopy. As shown in figure 3.23A, the 38-67°C UV hypochromic transition corresponding to the G quadruplex dissociation in the free RNA is shifted in the range 50-72°C when the RNA is complexed with the FXR1 RGG box, corresponding to a T_m of ~61°C (table 3.5), suggesting that the FXR1 RGG box stabilizes the G quadruplex structure of S3F-M2 RNA. Similar to the FMRP RGG box, the FXR1 RGG Box induces the formation of an alternate G quadruplex structure in S3F-M2 RNA upon binding. Analysis of the FXR2 RG cluster upon the G quadruplex structure of S3F-M2 RNA showed that this peptide has no effect upon the stability of the RNA (table 3.5) and it does not promote the formation of a secondary alternate structure (figure 3.23B)

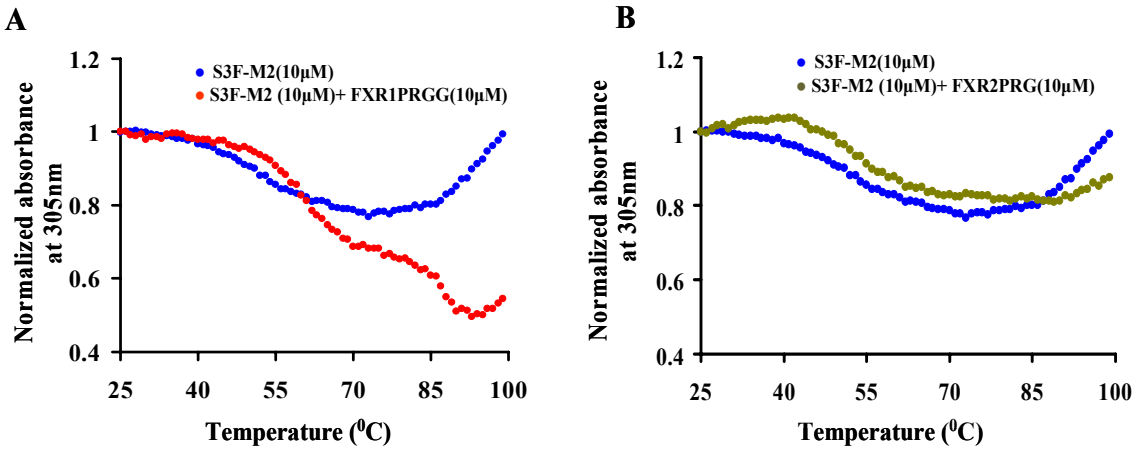

Figure 3.23 (A) UV melting profile of free 10 μM S3F-M2 RNA (blue trace) and in a 1:1 ratio with the FXR1P RGG box (red trace). (B) UV melting profile of free 10 μM S3F-M2 RNA (blue trace) and in a 1:1 ratio with the FXR2P RG (light green trace).

Molecule	T_m(°C)	$\Delta H°_{VH}$ (kcal/mol)	$\Delta S°_{VH}$ ($calmol^{-1}K^{-1}$)	$\Delta G°_{VH}$ at 25°C (kcal/mol)
S3F-M2	51.6 ± 0 .1	-43.1 ± 0.1	-132.1 ± 0.4	-3.6 ± 0.1
S3F-M2 + FXR1P RGG	60.6 ± 0.1	-50.7 ± 0.1	-152.1 ± 0.4	-5.4 ± 0.1
S3F-M2 + FXR2P RG	53.0 ± 0.2	-39.7 ± 0.1	-121.6 ± 0.7	-3.4 ± 0.1

Table 3.5. Thermodynamic parameters for the S3F-M2 RNA G quartet formation in the presence of FXR1P RGG Box and the FXR2 RG cluster.

3.6 The FMRP and FXR1P RGG box binding in superstoichiometric amounts unwind the S3F-M2 G quadruplex structure.

To investigate the effect of the FMRP and FXR1P RGG boxes on the intramolecular G quartet structure of S3F-M2 RNA, we compared the CD spectra of the free RNA with those of the RNA in complex with the RGG peptides. At a 1:1 ratio of S3F-M2 RNA: RGG peptide,the intensity of the CD band at 263-238 nm was almost unchanged, for both the FMRP and FXR1P RGG boxes (figure 3.24 A and B). Both peptides induced the unstacking of the G qudruplex structure of S3F-M2, as reflected by a decrease of intensity of the 263 nm CD band as well as a shift to 265 nm at higher ratios of the RNA: RGG box. We also performed the same experiment in the presence of the non-binding FXR2 RG cluster as a negative control (figure 3.25).

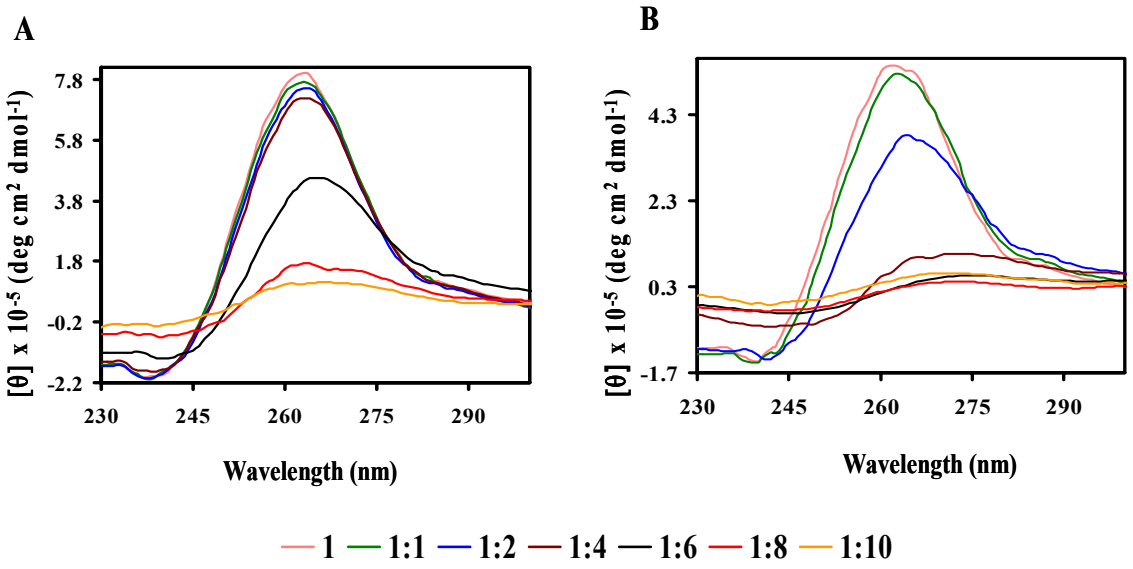

Figure 3.24 CD spectra of 10 µM S3F-M2 RNA in the presence of increasing concentrations of the RGG boxes of FMRP **(A**) and FXR1P **(B).**

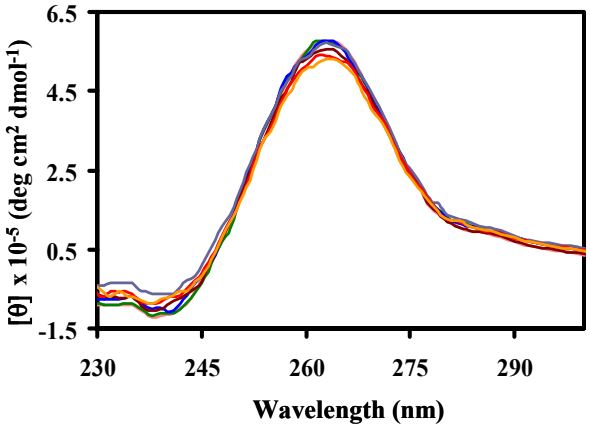

Figure 3.25 CD spectra of 10 µM S3F-M2 RNA in the presence of increasing concentrations of the FXR2P RG.

It is interesting to note that the FXR1P RGG peptide starts to unwind the G quadruplex RNA structure at only 1:2 RNA: peptide ratio, whereas this effect occurs for the FMRP RGG box at a 1:6 ratio. The FXR2P RG cluster does not seem to have any effect on the G quadruplex structure of S3F-M2 RNA.

In a control experiment we treated the solution containing the 1:10 RNA: peptide complexes with proteinase K, which degrades the RGG peptides, and re-acquired its CD spectrum (corrected for the proteinase K contribution) to rule out the possibility that the RNA is degraded in the presence of the large excess of the RGG peptides. The removal of the RGG peptides, allows the free RNA to refold into a G quadruplex structure, indicating that the spectral changes we observed in the presence of a large excess of the RGG peptides are not due to RNA degradation and that the unwinding we observe in the presence of the peptides is due to the G quadruplex structure unstacking.

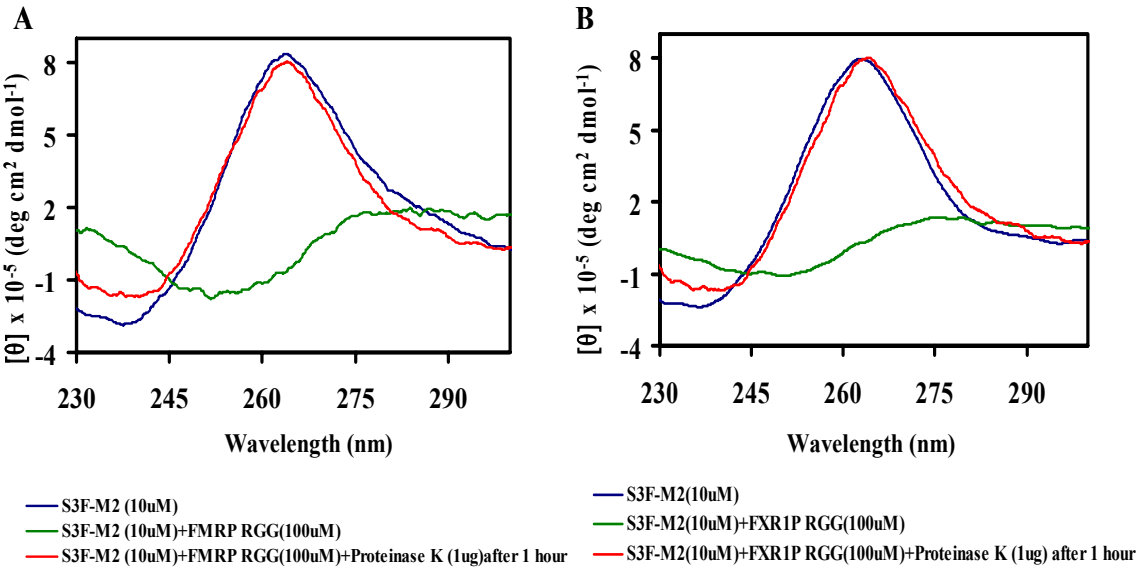

Figure 3.26 (A) 10 µM S3F-M2 RNA+ 100 µM FMRP RGG Box and **(B)** the 10 µM S3F-M2 RNA+100 µM FXR1P RGG Box with proteinase K (1µg) for 1 hour at 25^0C, to check for the degradation of the RNA.

Chapter 4

4.Microtubule Associated Protein 1B mRNA

4.1 The Microtubule Associated Protein 1B RNA forms an intramolecular G quadruplex structure.

In 2001, Darnell *et al*. proposed that a sequence rich in guanine residues located in the MAP1B mRNA 5'-untranslated region could fold into a G quadruplex stem-capped structure that is recognized by the FMRP RGG box (44). In 2003, Ramos *et al*. used NMR spectroscopy to demonstrate that a 43-mer MAP1B RNA oligonucleotide containing this guanine rich sequence folds into a G quadruplex structure which is recognized by the FMRP RGG box. Their observations, however suggested that heterogeneity exists in the binding of the RGG peptide to this RNA, since the MAP1B RNA existed in different conformations (45). In our studies we used a 34-mer MAP1B RNA fragment which retains the guanine-rich sequence proposed to form a G quadruplex stem capped structure. We introduced a point mutation at position 34 (G→C) to allow for the formation of a perfect seven base pair stem (figure 4.1). This mutation was not expected to affect the FMRP binding event, since it was shown that although a stem is required, its sequence is not important for FMRP binding to G quadruplex forming RNA (44).

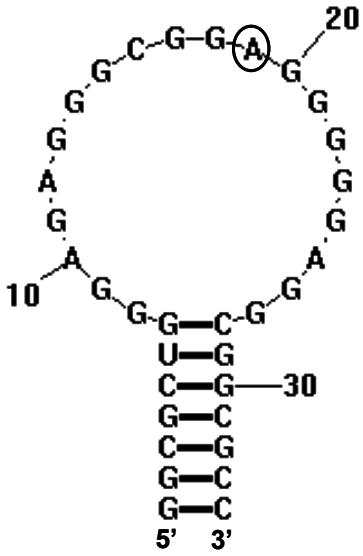

Figure 4.1 Secondary structure of the MAP1B RNA fragment (34-mer) used in this study, which was generated using Zuker's algorithm, 2003 (83).

We investigated if the 34-mer MAP1B RNA still maintains the ability to form a G quadruplex structure, by 1D ^{1}H NMR spectroscopy, focusing on the spectral changes in the imino proton region (figure 4.2). At 0 mM KCl, we observed only resonances corresponding to Watson-Crick imino protons, confirming the presence of a stem structure in MAP1B RNA. As the KCl concentration was increased to 4 mM a new set of resonances appeared in the region 10.8-12 ppm, corresponding to imino protons of guanines involved in G quartet formation, indicating that the 34-mer MAP1B RNA fragment forms a G quadruplex structure. Samantha Mader, in her preliminary experiments in our lab (personal communication), used 1D 1H NMR spectroscopy to demonstrate that MAP1B RNA is in slow exchange between two different conformations at KCl concentrations higher than 10 mM. Thus, to ensure the formation of a single

MAP1B RNA conformer, all further experiments were performed in the presence of 10 mM KCl.

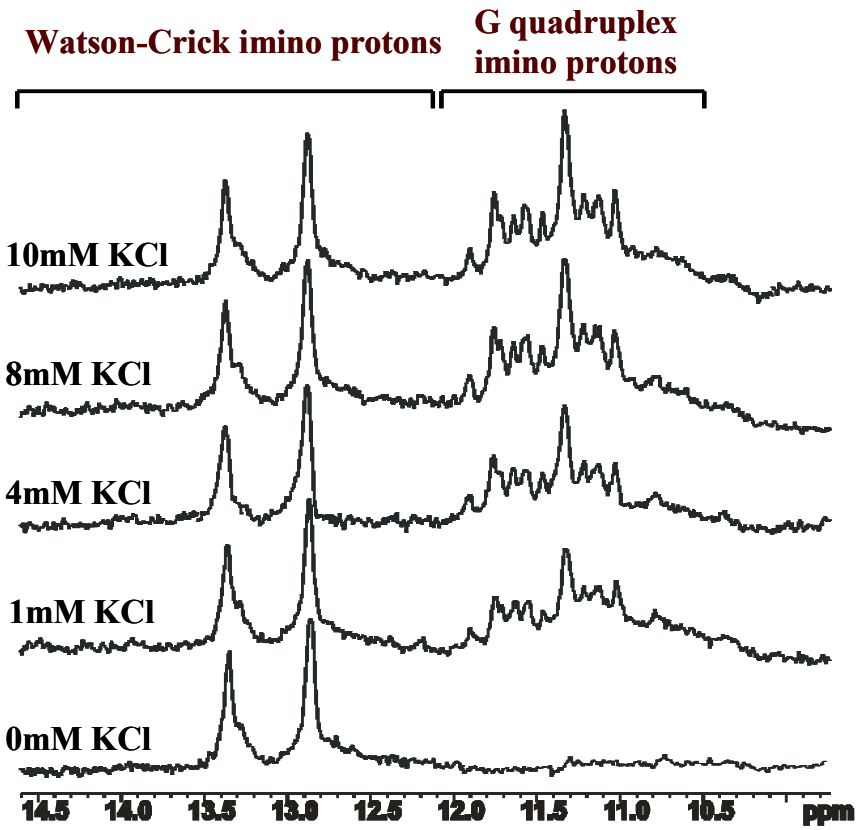

Figure 4.2 Changes of the imino proton resonance region of the 1D ^{1}H NMR spectrum of MAP1B RNA (500 μM in 10 mM Tris, pH 7.5, 29°C) upon titrating increasing concentrations of KCl in the range of 0-10 mM.

To determine if the G quadruplex fold of MAP1B RNA is *parallel* or *anti-parallel*, we monitored the spectral changes in the CD spectra of MAP1B RNA (10 μM in 10 mM cacodylic acid, pH 6.5) upon the titration of increasing KCl concentrations in

the range 0 - 10 mM. A positive band at 265 nm and a negative band at 240 nm were observed, corresponding to a type I G quadruplex CD spectrum (88, 89). Even though there are not enough intramolecular parallel type I high-resolution G quadruplex structures available, some examples have been reported (90-93). The CD spectrum of MAP1B RNA also shows a positive shoulder around 290 nm, a spectral feature that has been previously reported for intramolecular G-quadruplexes with *anti*-glycosidic bonds and external or "propeller" loops (94, 95). Alternately, this 290 nm shoulder could originate from a minor G quadruplex conformation present in MAP1B RNA (indicated by a star in the native gel electrophoresis gel shown in figure 4.19), which could be of anti-parallel nature, showing thus a type II CD spectrum with a positive band at 295 nm and a negative one at 260 nm (figure 4.3).

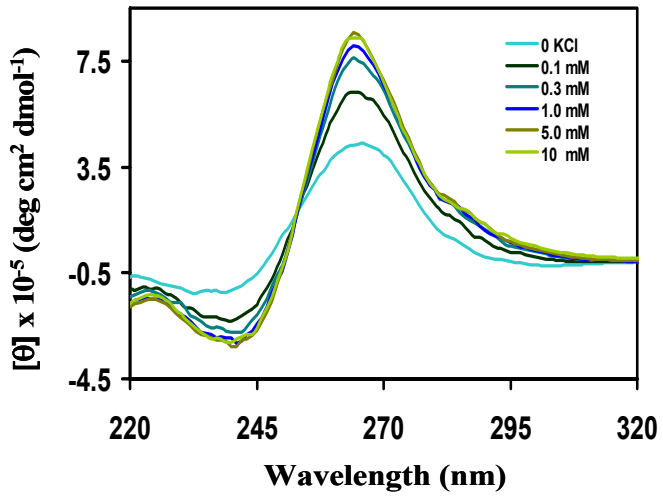

Figure 4.3 CD spectra showing the type I G quadruplex formation by MAP1B RNA (10 µM RNA in 10 mM cacodylic acid, pH 6.5 in the presence of increasing concentrations of KCl ranging from 0 - 10 mM). The KCl concentrations are indicated below the figure.

To investigate, whether the fold in the G quadruplex structure in the MAP1B RNA is *intramolecular* or *intermolecular* in nature, we used thermodynamic methods. It has been shown that for an intermolecular species with *n* number of strands the melting point of the G quadruplex structure, T_m, varies with the total RNA concentration (see equation 1 in materials & methods), whereas for an intramolecular species (*n=1*), the T_m does not depend on the RNA concentration (see equation 2 in materials & methods). First we measured the heating (20-99°C) and cooling (99-20°C) curves of MAP1B RNA by UV spectroscopy at 295 nm (57). The curves were super-imposable suggesting that the heating rate of 0.2°C/minute used in all thermal denaturation experiments allowed for the measurement of true equilibrium curves in MAP1B RNA (figure 4.4 A). A hypochromic transition with a T_m ~ 67ºC was observed in the UV thermal denaturation profile of MAP1B RNA in the presence of K^+, which we attributed to the dissociation of a G quadruplex structure (78). The hypochromic transition seen in the presence of K^+ was not observed in the thermal denaturation profile of MAP1B RNA folded in the presence of Li^+ ions, since typically these ions do not support the formation of a G quadruplex structure (figure 4.4 B).A hyperchromic transition starting around 77°C is also observed in the UV melting curve of MAP1B RNA, and initially we attributed this transition to the melting of the MAP1B RNA stem structure. This hypothesis however, was not supported when we monitored the imino proton resonance region of the MAP1B RNA 1D ^{1}H NMR spectrum, (where the stem and quadruplex imino protons give rise to distinct resonances), as a function of temperature (figure 4.5). The Watson-Crick imino proton resonances corresponding to the stem structure of MAP1B RNA completely disappeared above

65°C, suggesting that the UV hyperchromic transition starting at 77°C does not originate from the melting of the stem structure.

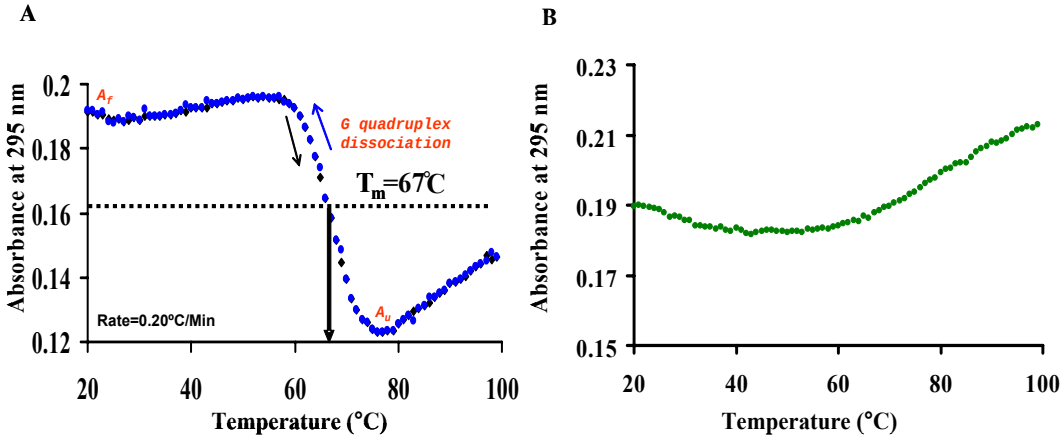

Figure 4.4 (A) The heating and cooling UV thermal denaturation profiles of MAP1B RNA (10μM in 10 mM cacodylic acid, pH 6.5 and 10 mM KCl) (black diamonds 20-99°C, open triangles 99-20°C).**(B)** UV thermal denaturation profile of MAP1B RNA (10 μM) in 10 mM cacodylic acid, pH 6.5) in the presence of 10 mM LiCl.

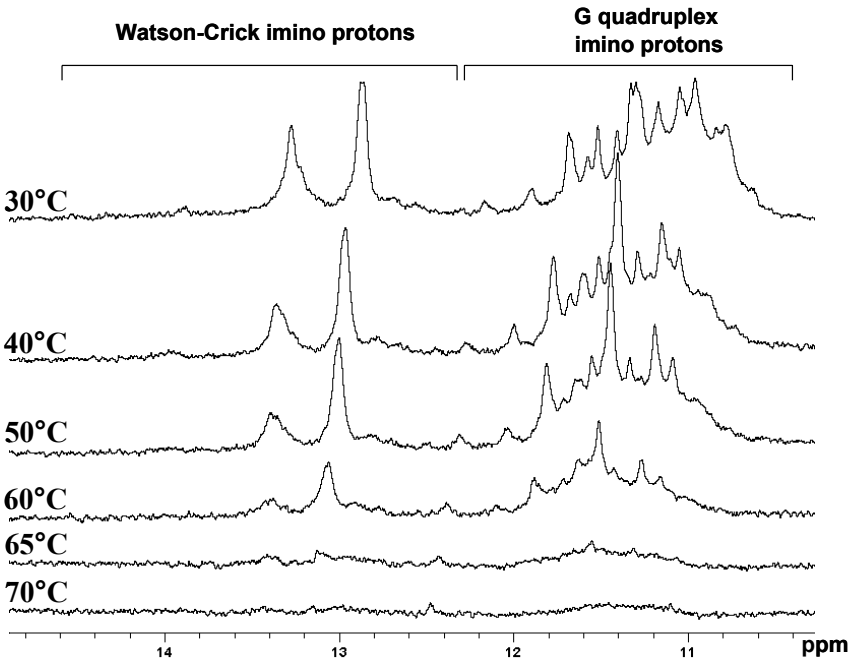

Figure 4.5 1D ^{1}H NMR spectrum showing the thermal denaturation profile of MAP1B RNA (365 μM MAP1B RNA in 10 mM Tris, pH 7.5 and 10 mM KCl) in the range 30°- 70°C.

We propose that, since the G quadruplex forming sequence of MAP1B RNA contains a stretch of twenty purines (starting at G7) interrupted by a single C, it is possible that upon the melting of the quadruplex structure, the liberated rG residues can stack on their nearest rA neighbors and the melting of these rG-rA stacks could give rise to the observed UV hyperchromic transition (47, 85, 86).

The melting temperature of the MAP1B RNA G quadruplex structure was measured at various RNA concentrations in the range 3-100 μM. The T_m remained constant at all RNA concentrations (figure 4.6), indicating the formation of an *intramolecular* G quadruplex structure in MAP1B RNA (see equation 2 in materials & methods). This result was also confirmed by native gel electrophoresis (data not shown).

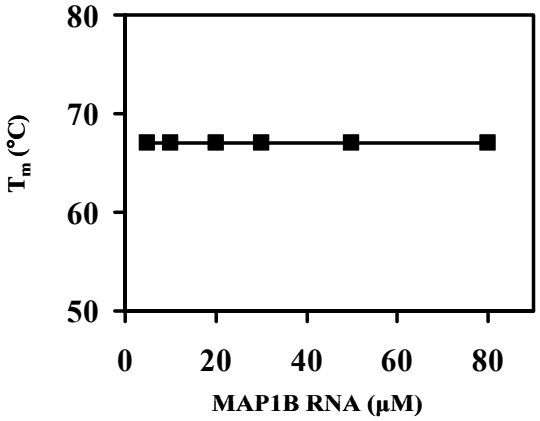

Figure 4.6 Plot of the MAP1B RNA G quadruplex melting temperature as a function of the RNA concentration.

The thermodynamic parameters for the MAP1B G quadruplex formation were obtained by fitting the UV hypochromic transition in the range 56-79ºC observed in the presence of K^+ with equation 3 (materials & methods). The values of the enthalpy and free energy of G quadruplex formation (ΔH^0 = -96.9 ± 0.1 kcal/mol and ΔG^0 = -12.1 ± 0.1 kcal/mol) suggest that the structure of MAP1B RNA contains four to five G quartet planes (the enthalpy of formation of a single G quartet plane in an intramolecular G quadruplex, measured in similar experimental conditions, ranges from -18 to -25 kcal/mol, (87). However, the MAP1B RNA sequence does not seem to contain enough uninterrupted four guanine stretches (figure 4.1) to fold in to a structure containing four to five G quartet planes. We therefore propose that this structure might contain mixed purine (adenine and guanine) quartets. These structural elements have also been proposed to exist in the FMR1 mRNA and bind to the FMRP specifically and with high affinity (43).

Next, we determined the number of specifically bound K^+ ion equivalents in the MAP1B RNA G quadruplex structure, by measuring the T_m of the RNA at different KCl concentrations ranging from 0.1 mM-10 mM. A simple model for the folded to unfolded G quadruplex MAP1B RNA structure was assumed, in which Δn K^+ ion equivalents are released upon unfolding. Δn can be determined from the slope of the plot of ΔG^0 as a function of logarithm of K^+ ion concentration (96):

$$\Delta n = \frac{d \ln K_{eq}}{d \ln [K^+]} = -\frac{\Delta\Delta G^0}{2.3RT\Delta \log [K^+]} \qquad (9)$$

where $\ln K_{eq} = -\Delta G^0/RT$ and $\Delta\Delta G^0 / \Delta \log [K^+]$ is the slope of the plot of ΔG^0 as a function of logarithm of K^+ ion concentration (figure 4.7). A Δn value of 3.6 ± 0.2, was

obtained, indicating that the MAP1B G quadruplex structure is coordinated by approximately four K^+ ion equivalents.

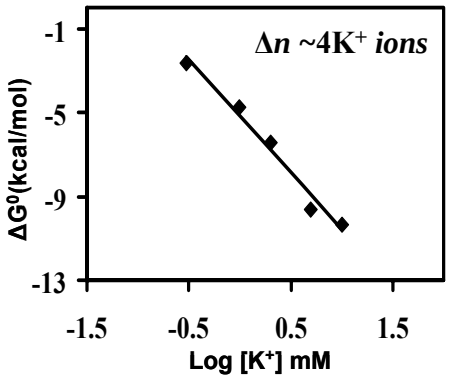

Figure 4.7 Plot of ΔG^0 as a function of the logarithm of K^+ ion concentration. The number *Δn* of K^+ ion equivalents released upon the unfolding of MAP1B RNA G quadruplex structure was determined from the slope $\Delta\Delta G^0 / \Delta \log [K^+]$.

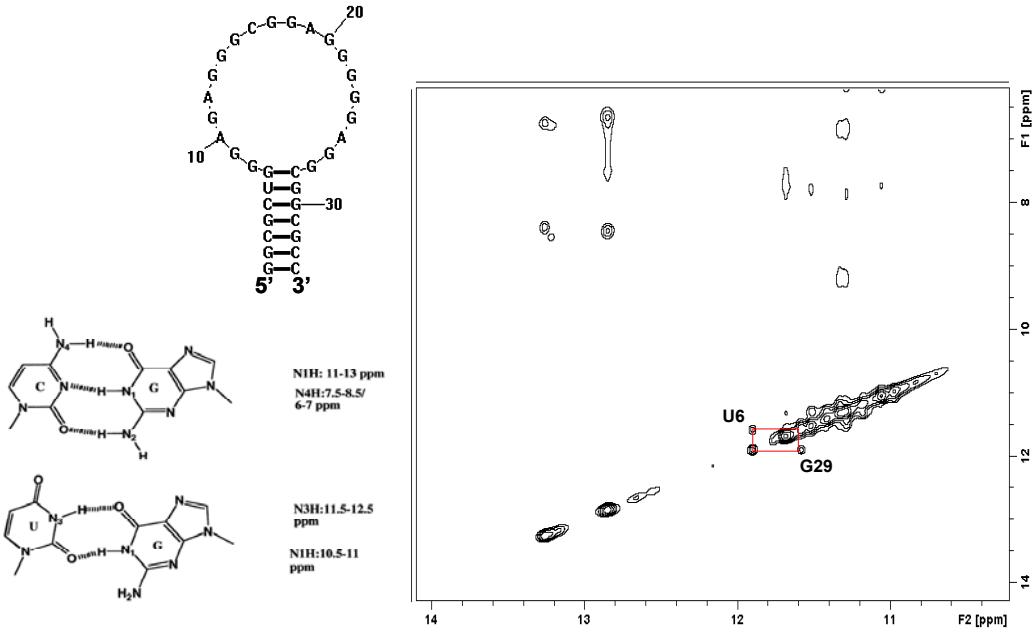

Figure 4.8 ^{1}H-^{1}H NOESY NMR spectrum of MAP1B RNA (300 μM) in 10 mM Tris, pH 7.5 containing 10 mM KCl .

We also performed a 2D ^{1}H-^{1}H NOESY NMR experiment on MAP1B RNA to obtain higher resolution information about its structure. The sample concentration was found to be too low to obtain a high quality 2D spectrum (figure 4.8). Nonetheless, we observe a strong NOE pattern that indicates the presence of a G-U base pair, which we assign to U6 and G29.

4.2 Interactions of the G quadruplex forming MAP1B RNA with the FMRP RGG box.

Our thermodynamic and CD spectroscopy results indicate that the 34-mer MAP1B RNA maintains the ability to form an intramolecular parallel G quadruplex structure, which contains four to five G quartet planes and is coordinated by ~ $4K^{+}$ ion equivalents. This structure needs to be confirmed by the high-resolution structure determination of this RNA.

Next, we performed a detailed thermodynamic study of the interactions between the G quadruplex forming MAP1B RNA and the FMRP RGG box, by using fluorescence spectroscopy. In these studies, the MAP1B RNA was labeled by the fluorescent analog 2 aminopurine (2AP) at position 19 (circled in figure 4.1), constructing MAP1B-19AP RNA.

In control experiments, we demonstrated that the 2AP insertion does not affect the secondary structure and stability of the MAP1B RNA by comparing the UV thermal profiles and the CD spectra of MAP1B and MAP1B-19AP RNA. As seen in figures 4.9 A and B, the UV thermal melting profiles and the CD spectra of both RNAs were similar.

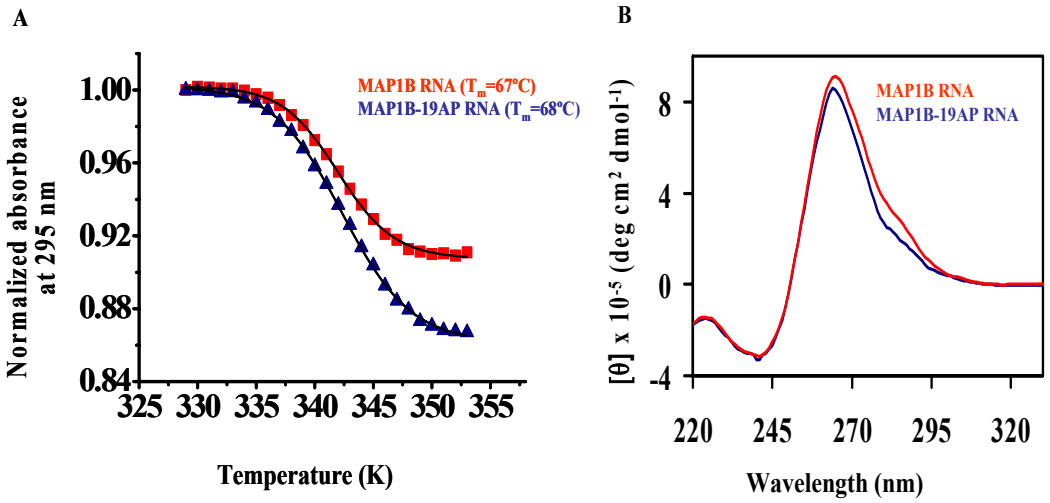

Figure 4.9 **(A)** UV melting profiles and **(B)** CD spectra of MAP1B and MAP1B-19AP in 10 mM cacodylic acid, pH 6.5 and 10 mM KCl.

We also determined by electromobility gel shift assay that the 2AP insertion in MAP1B-19AP did not affect its ability to bind to the FMRP RGG box (figure 4.10).

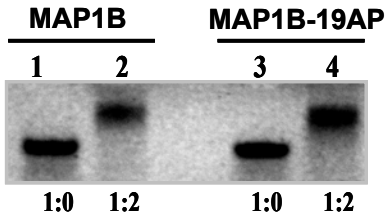

Figure 4.10 Electrophoretic mobility gel shift analysis of the FMRP RGG box interactions with MAP1B and MAP1B-19AP RNA. The RNA concentration was 10 μM for both MAP1B (lane 1) and MAP1B-19AP RNA (lane 3).The FMRP RGG box added in a 1:2 RNA: peptide ratio shifted MAP1B (lane 2) and MAP1B-19AP RNA (lane 4) in a similar manner. Both the gel and running buffer contained 10 mM KCl. The gel was visualized by UV shadowing at 254 nm.

We hypothesized that the environment of the 2-AP reporter will be different when MAP1B RNA is folded in the presence of Li^+ ions that do not promote G quadruplex formation versus K^+ ions (single stranded loop versus a G quadruplex-surrounding loop), resulting in a difference in its steady-state fluorescence, since the quantum yield of 2AP is very sensitive to its microenvironment. A two-fold change in the 2AP steady state fluorescence was indeed observed when MAP1B-19AP RNA was folded in the presence of K^+ ions, as opposed to Li^+ (figure 4.11A). We also performed a control experiment using a 16 nt RNA named DLS_9AP, which has a 2AP incorporated between two guanines (its sequence is described in the materials and methods section), but which cannot form a G quadruplex structure. We observed that the steady-state fluorescence of the 2AP incorporated in DLS_9AP is the same in the presence of 10 mM LiCl and 10 mM KCl (figure 4.11B), indicating that the observed changes in the 2AP steady state fluorescence are due to the formation of the MAP1B G quadruplex structure, and not due to an effect of the different ions.

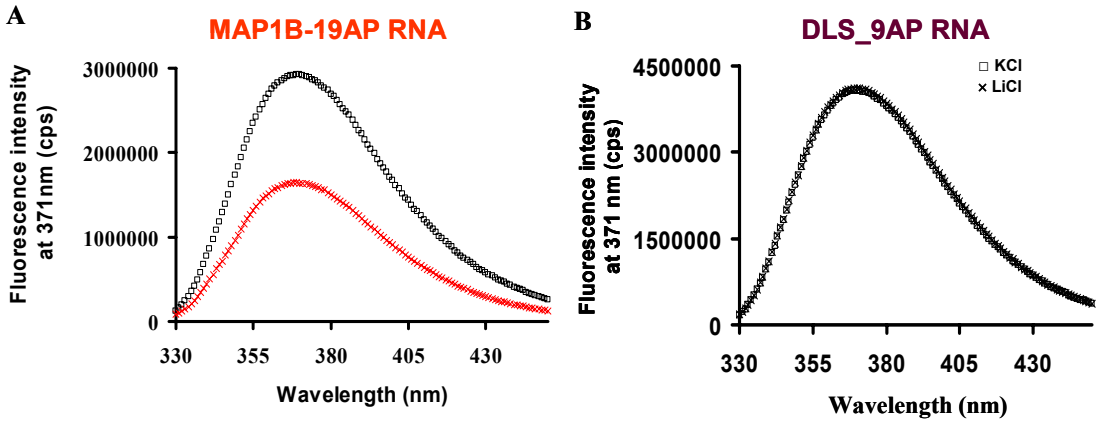

Figure 4.11 (A) The 2-AP at the 19th position reports on the G quadruplex structure formation, as indicated by the change in its steady state fluorescence when MAP1B-19AP RNA is folded in the presence of KCl (squares) as opposed to LiCl (plus sign). **(B)** The presence of different cations (Li^+ versus K^+) does not affect the steady-state fluorescence intensity of a 2-AP reporter inserted in the non-quadruplex forming DLS_9AP RNA.

The binding of the FMRP RGG box to the MAP1B-19AP RNA was measured by monitoring the 2AP steady state fluorescence changes upon titrating increasing amounts of the FMRP RGG peptide (figure 4.12, black squares) into a fixed concentration of MAP1B RNA (400 nM). The dissociation constant for the MAP1B RNA: FMRP RGG complex, determined by fitting the binding curve with equation 4 (materials & methods), was found to be $K_d = 20.1 \pm 6.4$ nM, corresponding to a free energy of binding $\Delta G^0_{obs} = -10.4 \pm 0.2$ kcal/mol. This K_d is significantly lower than the value reported by Darnell *et al*. ($K_d = 500$ nM) for the binding of either the full-length FMRP or of the isolated RGG box domain to MAP1B RNA (44). One possible explanation for this discrepancy could be the different experimental conditions of the binding assay. Darnell reports that their binding measurements were performed in the presence of 200 mM K^+, conditions in which we show that MAP1B RNA exists in more than one conformation. Thus, the 500 nM value is likely an average for the binding of FMRP to these different MAP1B RNA conformers. Our binding experiments performed under conditions where MAP1B exists mainly as a single conformer in the presence of 10 mM K^+.

To determine if the FMRP RGG box binds specifically to MAP1B RNA, we measured the K_d values of the MAP1B RNA: FMRP RGG complex in the presence of a 10-fold excess of non specific RNAs, and in the presence of a 10-fold excess of the non-specific peptide HCV peptide (sequence of this peptide is listed in the materials and methods). We demonstrate that the presence of a 10-fold excess of the HCV core peptide does not affect the binding of the FMRP RGG box to MAP1B RNA since the dissociation constant measured in these conditions is $K_d = 18.7 \pm 6.7$ nM (figure 4.12 (black crosses). The dissociation constant values of the MAP1B RNA-FMRP RGG

complex were also measured in the presence of a 10-fold excess of two non-specific RNAS: Munc 13 site 1 RNA and rComp4 RNA, which are described in the materials and methods section. The K_d values measured in the presence of an excess of either RNA (K_d = 23.0 ± 3.8 nM for Munc13 site 1 RNA and K_d = 19.0 ± 8.6 nM for rComp4 RNA) were found to be similar to those measured in the absence of the non-specific RNAs. These results indicated that the FMRP RGG box binds specifically and with high affinity to the MAP1B RNA G quadruplex structure.

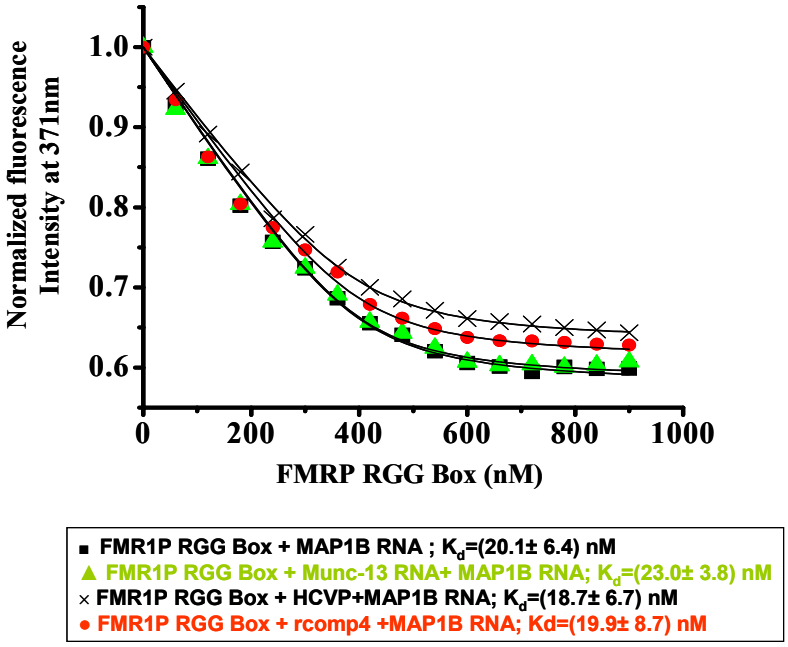

Figure 4.12 Binding curves of the FMRP RGG box to MAP1B-19AP RNA in the absence (black squares) and presence of a 10-fold excess of a non-specific Munc-13 site 1 RNA (black triangles) or of a nonspecific HCV peptide (black crosses) or of rcomp4 sequence (black circles). These experiments were carried out at 25°C.

Since the sequence of Munc 13 site 1 RNA is rich in Gs, we inquired if it is able to form a G quadruplex structure. Thus, we monitored the 10-11.5 ppm region of its 1D ^{1}H NMR spectrum and found that indeed this RNA adopts a G quadruplex structure in the presence of 10 mM K^{+}, as evidenced by the presence of tetrad guanine imino proton resonances (figure 4.13A). This result was confirmed by CD spectroscopy, since the CD spectrum of the Munc 13 site 1 RNA revealed a type I parallel G quadruplex fold signature (figure 4.13B). The finding that the dissociation constants for the FMRP RGG box binding to the MAP1B RNA in the presence a 10-fold excess and the absence of Munc 13 site 1 RNA are similar suggests that the FMRP RGG box is able to distinguish among different type I parallel fold G quadruplex structures. It is not known if the FMRP recognition of its G quadruplex forming mRNA targets involves a structure-specific mechanism (FMRP would recognize the unusual G quadruplex "shape" of the RNA) or a sequence-specific mechanism (the G quadruplex would function to properly expose the nucleotides located in its surrounding loops to FMRP) or a combination of both mechanisms. Although both have a type I parallel fold, our finding that FMRP RGG box binds with high affinity to the MAP1B RNA G quadruplex, but not to the Munc 13 site 1 RNA G quadruplex, indirectly suggest a sequence-specific mechanism of recognition. However, only the high-resolution structure of the FMRP RGG box-MAP1B RNA complex will confirm the molecular details of their association.

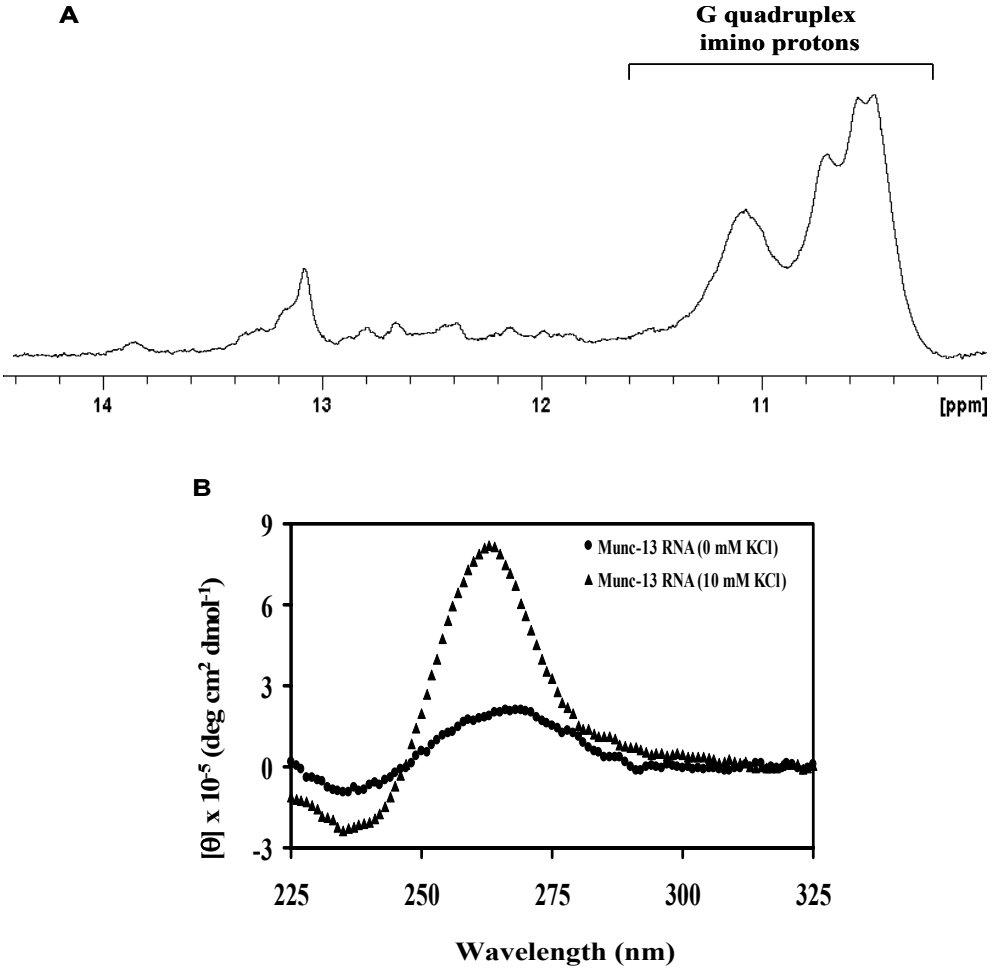

Figure 4.13 **(A)** Imino proton resonance region of the 1D ^{1}H NMR spectrum of Munc 13 site 1 RNA (263μM) in 10 mM Tris, pH 7.5 and 10 mM KCl. **(B)** CD spectra showing the type I G quadruplex formation by Munc 13 site 1 RNA (10 μM RNA in 10 mM cacodylic acid, pH 6.5 in the absence and presence of 10 mM KCl, respectively).

We tested in a control experiment whether Munc-13 site 1 RNA competes with the G quadruplex structure of MAP1B RNA for the FMRP RGG box or determine if Munc-13 site 1 RNA can displace the MAP1B RNA. First, we titrated an excess of the FMRP RGG box to a fixed concentration of MAP1B RNA until it reached saturation, and then titrated increasing concentrations of the Munc-13 site 1 RNA, up to 10-fold excess over the MAP1B RNA concentration. The steady-state fluorescence of the 2-AP did not change indicating that the FMRP RGG box binds specifically and with high affinity to the MAP1B RNA (figure 4.14).

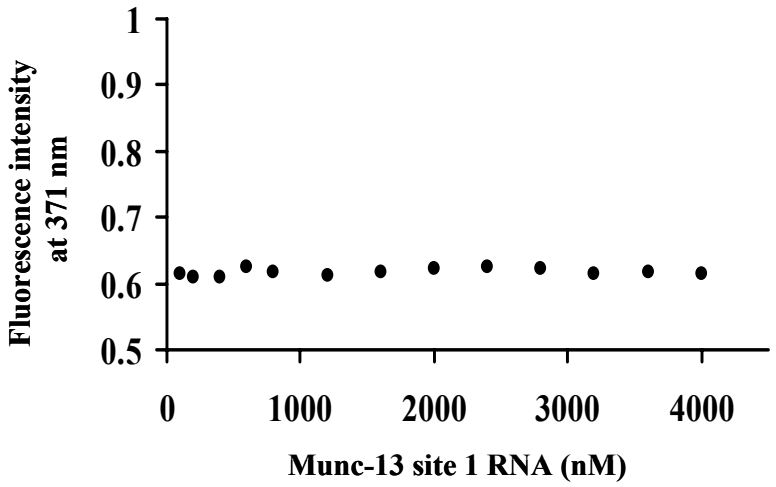

Figure 4.14 Increasing concentrations of Munc-13 site 1 RNA were titrated in to the MAP1B RNA-FMRP RGG Box complex.

Next, to obtain a more complete description of the forces driving the association between MAP1B-19AP RNA and the FMRP RGG box, we have determined the enthalpic and entropic contributions to the free energy of binding. The equilibrium association constant, $K_{obs} = 1/K_{d,}$ was measured as a function of temperature, in the range 20-45ºC. When the standard enthalpy, ΔH^0_{obs} and entropy, ΔS^0_{obs} do not depend on temperature, the van't Hoff plot which shows the dependence of $\ln(K_{obs})$ upon 1/T is linear. However, the van't Hoff plot for the FMRP RGG box binding to MAP1B RNA is non-linear (table 4.1) , indicating that ΔH^0_{obs} and ΔS^0_{obs} depend on temperature (figure 4.15 A).

Temperature (^{0}C)	K_d (nM)
20	21.0 ± 5.8
25	20.6 ± 6.5
30	22.5 ± 3.3
35	24.3 ± 3.7
40	27.4 ± 8.2
42	56.3 ± 14.0
45	62.4 ± 22.0

Table 4.1 The binding association constant was measured, $K_{obs} = 1/K_d$ for the MAP1B-19AP RNA - FMRP RGG box complex, as a function of temperature, in the range 20-45ºC.

To determine the entropy and enthalpy changes at different temperatures, the non-linear van't Hoff plot was fit with equation 8 (materials and methods). The heat capacity change of the system, $\Delta C^0_{p\,obs}$, and the temperatures, T_H and T_S, at which the enthalpy and entropy do not make any contributions to the energy of the system were obtained from this fit. The values, $\Delta C^0_{p\,obs}$ = -1.0 ± 0.2 kcal/mol K, T_H = 296 ± 2 K (23°C) and T_S = 307 ± 1 K (34°C) were then used in equations 6 and 7 (materials and methods) to obtain ΔH^0_{obs} and ΔS^0_{obs} at various temperatures. The standard entropy and enthalpy changes vary drastically with the temperature, while the binding free energy remains constant (figure 4.15B). This entropy-enthalpy compensation indicates a change in the nature of forces driving the complex formation: below 23°C entropy-driven ($\Delta S^0_{obs}>0$) and enthalpy-opposed ($\Delta H^0_{obs}>0$), between 23°C and 34°C entropy- and enthalpy-driven ($\Delta H^0_{obs}<0$ and $\Delta S^0_{obs}>0$), and above 34°C enthalpy-driven ($\Delta H^0_{obs}<0$) and entropy-opposed ($\Delta S^0_{obs}<0$). The FMRP RGG box-MAP1B RNA association at the physiological temperature is enthalpically driven with an unfavorable entropic contribution. This phenomenon has also been reported for the G quadruplex forming RNA targets of the FMRP RGG Box namely the Sc1 RNA (46) and semaphorin 3F RNA (47).

Hydrophobic interactions are likely significant in the FMRP RGG box-MAP1B RNA association, as indicated by the large negative value of $\Delta C^0_{p\,obs}$ = -1.0 ± 0.2 kcal/mol K. $\Delta C^0_{p\,obs}$ has been related to the change in polar (ΔA_p) and non-polar (ΔA_{np}) surface area (Å^2) occurring during complex formation (97):

$$\Delta C^0_{p\,obs} = 0.32\,\Delta A_{np} - 0.14\,\Delta A_p \qquad (10)$$

A large negative value for $\Delta C^{0}_{p\,obs}$ can be obtained from a large ($-\Delta A_{np}$) value, which results from hydrophobic interactions that bury the nonpolar surface of the forming complex and reduce exposure to water (81, 97, 98). However, in this process the expulsion of ordered water molecules from the binding surfaces typically contribute favorable changes in entropy (97). The finding that unfavorable changes of ΔS^{0}_{obs} are observed for the FMRP RGG box: MAP1B RNA complex formation at the physiological temperature indicates that structural changes in the peptide and/or RNA are likely associated with the binding event. Such structural changes could result in the restriction of molecular motion, and/or could affect the ionic and/or hydration shells around the complex by enhancing the association of ions and additional water molecules with the complex. The sequence analysis of the FMRP RGG box (materials and methods) does not reveal obvious amino acid residues that could become involved in stacking interactions with the MAP1B RNA bases. One possible candidate is Phe549, or alternately, the guanidino group of some of the RGG box Arg residues might stack with bases located in the G quadruplex surrounding loops.

Next, we evaluated the role played by electrostatic interactions in the FMRP RGG box-MAP1B RNA complex formation, since the FMRP RGG box is rich in basic residues and the RNA is an anionic polymer. We measured the association constant of the complex, K_{obs}, in the presence of increasing cation [M^{+}] concentrations. The slope ($-\partial \log K_{obs}/\ \partial \log\ [M^{+}]$), known as the salt-dependence, is related to the number of electrostatic interactions that stabilize the peptide-RNA complex (99).

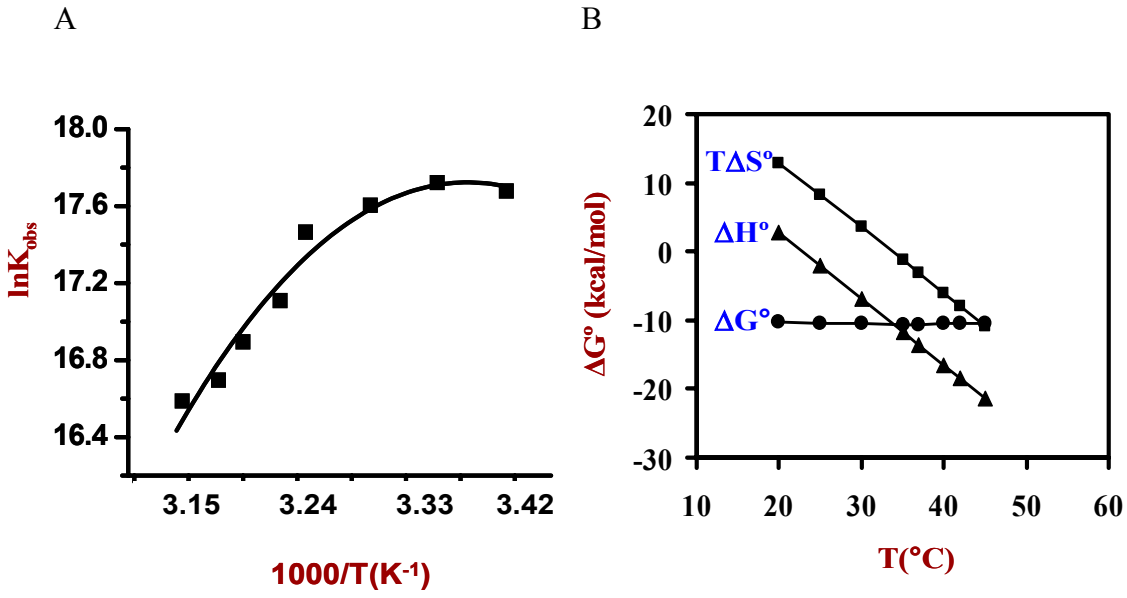

Figure 4.15 (A) Non linear van't Hoff plot, showing the temperature dependence of the association constant, $K_{obs} = 1/K_d$, for the MAP1B-19AP RNA: FMRP RGG complex, in the range of 20-45°C. The $\Delta C°_{pobs}$ = - (1.0 ± 0.2) kcal/ mol K, T_H (K) = (296 ± 2); T_S (K) = (307 ± 1) values were determined by fitting this plot with equation 8 **(B)** Dependence of the entropy, enthalpy and free energy of the MAP1B-19AP RNA: FMRP RGG complex formation on temperature.

Since MAP1B RNA forms alternate species, in the presence of K^+ concentrations above 10 mM, the cation concentration $[M^+]$ was varied in the range 10 to 1000 mM by the addition of increasing concentrations of Li^+ ions. A change in the dissociation constant, from 20 nM in the presence of 10 mM $[M^+]$ to 38 nM in the presence of 1000 mM $[M^+]$ was observed for the FMRP RGG box: MAP1B-19AP RNA complex. The slope of this plot corresponded to a value ($-\partial \log K_{obs}/ \partial \log [M^+]$) of ~0.2. In protein-RNA complexes, where the electrostatic interactions are not significant, values from 0 to 2 ($-\partial \log K_{obs}/ \partial \log [M^+]$) have been measured (99-102). The difference in the free energy $\Delta\Delta G^0_{obs}$ in the presence of 10 mM and 1000 mM $[M^+]$ was calculated to be ~ 0.4 kcal/mol, representing a mere 3% of the total binding free energy. We conclude thus, that

electrostatic interactions do not play a dominant role in the association between the FMRP RGG box and MAP1B RNA. The electrostatics have also been shown to have minor contributions in the binding of the FMRP RGG box to two other G quadruplex forming RNAs, Sc1 RNA and Semaphorin 3F RNA (46,47).

Total salt concentration (mM)	K_d (nM)
10	20.6 ± 6.5
75	24.5 ± 5.0
150	31.5 ± 7.5
400	34.6 ± 8.1
1000	34.7 ± 8.1

Table 4.2 Binding of FMRP to the MAP1B-19 AP was measured at different salt concentrations to determine the role of electrostatic interactions in the complex formation. Salt composition =10mM KCl + variable LiCl (mM)

We have also investigated if the FMRP RGG box binding affects the stability of the G quadruplex structure of MAP1B RNA, by measuring the melting temperature of the RNA G quadruplex structure in the MAP1B-FMRP RGG box complex, by UV spectroscopy. This melting temperature was ~71ºC, which corresponds to a free energy of G quadruplex formation of ΔG^0 = -13.3 ± 0.1 kcal/mol (figure 4.17). Thus, we conclude that the binding of the FMRP RGG box in a 1:1 ratio to MAP1B RNA induces a slight stabilization of the RNA G quadruplex structure.

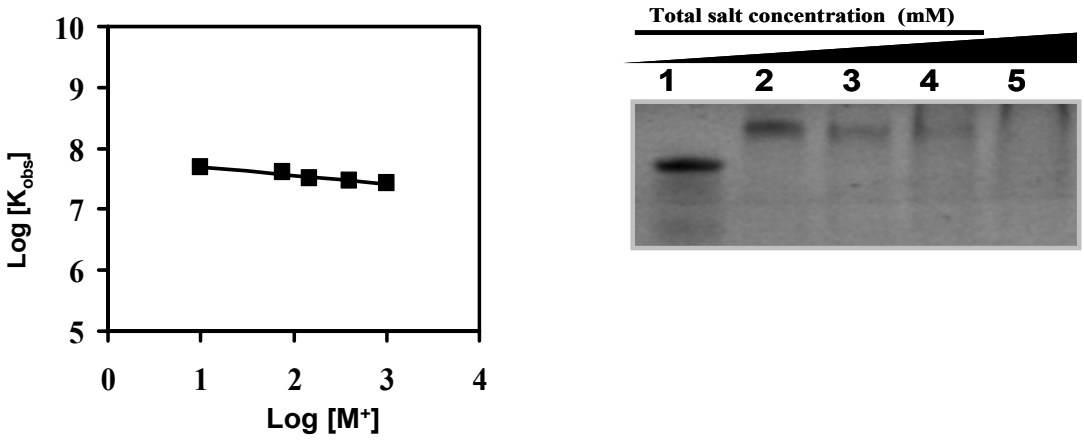

Figure 4.16 (A) The binding of FMRP RGG box to MAP1B-19AP was measured at 25°C in the presence of different salt concentrations [M^+] (Salt composition: 10 mM KCl + variable mM LiCl mM) in the range 10 mM to 1000 mM [M^+]. A value of ($-\partial \log K_{obs} / \partial \log[M^+]$) of ~0.2 was determined from the linear fit of the plot. **(B)** The results from the fluorescence experiments were also confirmed by EMSA. Lane 1 - MAP1B RNA (10 µM), Lane 2 - MAP1B RNA + FMRP RGG (10 mM KCl), Lane 3 - MAP1B RNA + FMRP RGG (150 mM KCl), Lane 4 -MAP1B RNA + FMRP RGG (400 mM KCl), Lane 5 - MAP1B RNA + FMRP RGG (1000 mM KCl).

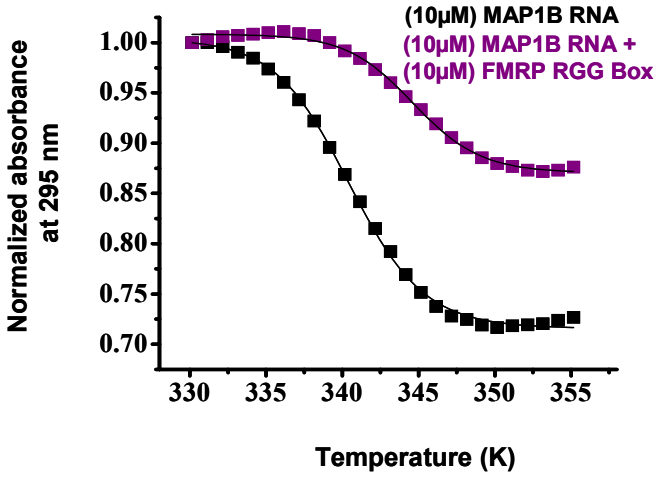

Figure 4.17 UV melting profile of MAP 1B RNA (black squares) and the MAP1B RNA + FMRP RGG Box (purple squares).

Molecule	Tm(^{0}C)	$\Delta H^0 f$ (kcal/mol)	$\Delta S^0 f$ (calmol^{-1}K^{-1})	$\Delta G^0 f$ (kcal/mol)
MAP1B	67.4 ± 0.1	-96.9 ± 0.1	-284.6 ± 0.2	- 12.1 ± 0.1
MAP1B+FMRP RGG	71.5 ± 0.1	-98.1 ± 0.1	-284.8 ± 0.2	- 13.3 ± 0.1

Table 4.3 Thermodynamic parameters for the binding of FMRP RGG Box to the MAP1B RNA.

However, our CD spectroscopy results showed that upon titrating an excess of the FMRP RGG box (RNA: peptide ratios higher than 1:8) a reverse effect is induced, the G quadruplex structure of the MAP1B RNA being unfolded (figure 4.18A). To ensure that the reduction of the 265 nm band intensity observed in the presence of an excess of the RGG box peptide was not due to the RNA degradation, we treated the sample containing a ratio of 1:10 MAP1B RNA-FMRP RGG peptide with proteinase K, which degrades the RGG box peptide. As seen in figure 4.18B, the MAP1B G quadruplex structure refolds upon the FMRP RGG peptide degradation, giving rise to the same CD spectrum as that of the free MAP1B RNA. We have also confirmed this result by NMR spectroscopy (figure 4.18C) upon increasing the FMRP RGG box concentration, the sharp resonances centered ~ 11.5 ppm (corresponding to guanine imino protons involved in Hoogsteen hydrogen bonds in the G quadruplex structure) disappear, indicative of the unfolding of the G quadruplex structure. Interestingly, the MAP1B RNA stem structure remains unaffected, suggesting that the interactions of the FMRP RGG box are limited to the G quadruplex structure. A similar effect of G quadruplex unfolding at high RNA: FMRP RGG peptide ratio was observed for the binding of the FMRP RGG box to the G quadruplex forming Semaphorin 3F RNA and Sc1 RNA (47 and data not shown). We have attempted to

speculate on a possible biological function for the G quadruplex RNA structure stabilization at low FMRP RGG box concentrations and its unwinding at high FMRP RGG box concentrations in the conclusions section of this study.

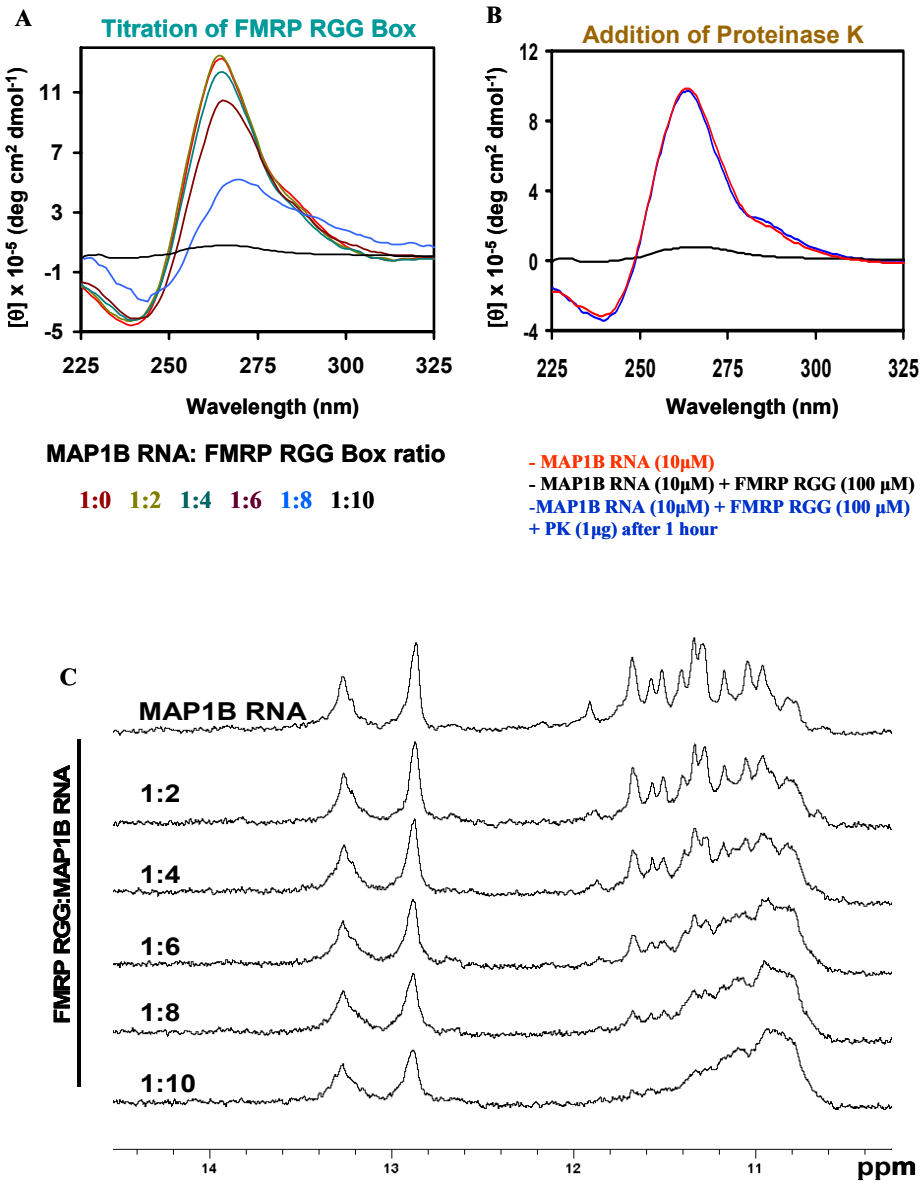

Figure 4.18 **(A)** CD spectra of 10 μM MAP1B RNA in the presence of increasing concentrations of the FMRP RGG box. **(B)** 10 μM MAP1B RNA+100 μM FMRP RGG Box was incubated with proteinase K (1μg) for 1 hour at 25^0C to check if the addition of an excess of the FMRP RGG box causes the RNA degradation. **(C)** Imino proton resonance region of the 1D ^{1}H NMR spectra of MAP1B RNA in the presence of increasing concentrations of the FMRP RGG box.

4.3 Interactions of the FMRP autosomal paralogs, FXR1P and FXR2P with MAP1B RNA.

To determine if the recognition of the MAP1B G quadruplex structure is unique to the FMRP RGG box or if it is a property shared by the FMRP autosomal paralogs FXR1P and FXR2P, we first performed gel electrophoresis shift assay. Our results indicated that the MAP1B RNA is bound by both the FXR1P RGG box and by the FXR2P RG cluster, respectively (figure 4.19, compare lanes 3 with 4 and lanes 5 with 6). The FXR1P has an RGG box domain and FXR2P has an RG cluster.

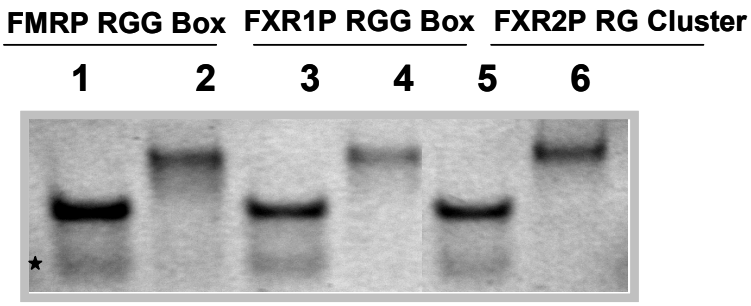

Figure 4.19 EMSA of the binding of MAP1B RNA by the FMRP RGG box (lanes 1 & 2), FXR1P RGG box (lanes 3 & 4) and FXR2P RG cluster (lanes 5 & 6).

Fluorescence spectroscopy was used to determine the binding curves of the FXR1P RGG and FXR2P RG cluster to MAP1B-19AP RNA. Increasing amounts of each peptide were added to a fixed amount of MAP1B-19AP RNA (figure 4.20A and B). The dissociation constants of the MAP1B RNA: FXR1P RGG box complex, K_d = 73.1 ± 10.1 nM, and of the MAP1B RNA: FXR2P RG cluster complex, K_d = 127.1 ± 45.0 nM, were determined by fitting the binding curves with equation 4 (materials & methods),

indicating that the MAP1B RNA is recognized by FXR1P and FXR2P with high affinity. However, among the fragile X protein family, FMRP has the highest affinity for this RNA (K_d = 20.1 ± 6.4 nM). The free energy of binding for the association between the FXR1 RGG box and the MAP1B RNA was determined to be ΔG^0 = -9.7 ± 0.2 kcal/mol. The free energy of binding between the FXR2 RG cluster and the MAP1B RNA was determined to be ΔG^0 = -9.4 ± 0.2 kcal/mol.

Next, we evaluated if the binding between the FXR1P RGG box and FXR2P RG cluster and the MAP1B-19AP RNA was specific in nature by measuring their binding curves in the presence of a 10-fold excess of either non-specific RNA (Munc13 RNA and rComp4 RNA), or of the non-specific HCV core peptide (figure 4.20A and B). The dissociation constants, obtained by fitting the binding curves with equation 4 (reported in the figure 4.20 legend) were found to be larger than those measured in the absence of the excess of non-specific RNA or peptide. The values remain in the nanomolar range, indicating that some specificity exists in the recognition of MAP1B-19AP by the FXR1P RGG box and the FXR2 RG cluster. Other G quadruplex forming RNA targets, such as Semaphorin 3F (47) and Sc-1 (46), have also been shown to be bound by FXR1P; however, to our knowledge this is the first report demonstrating that the FXR2P RG cluster binds to a G quadruplex forming RNA target of FMRP.

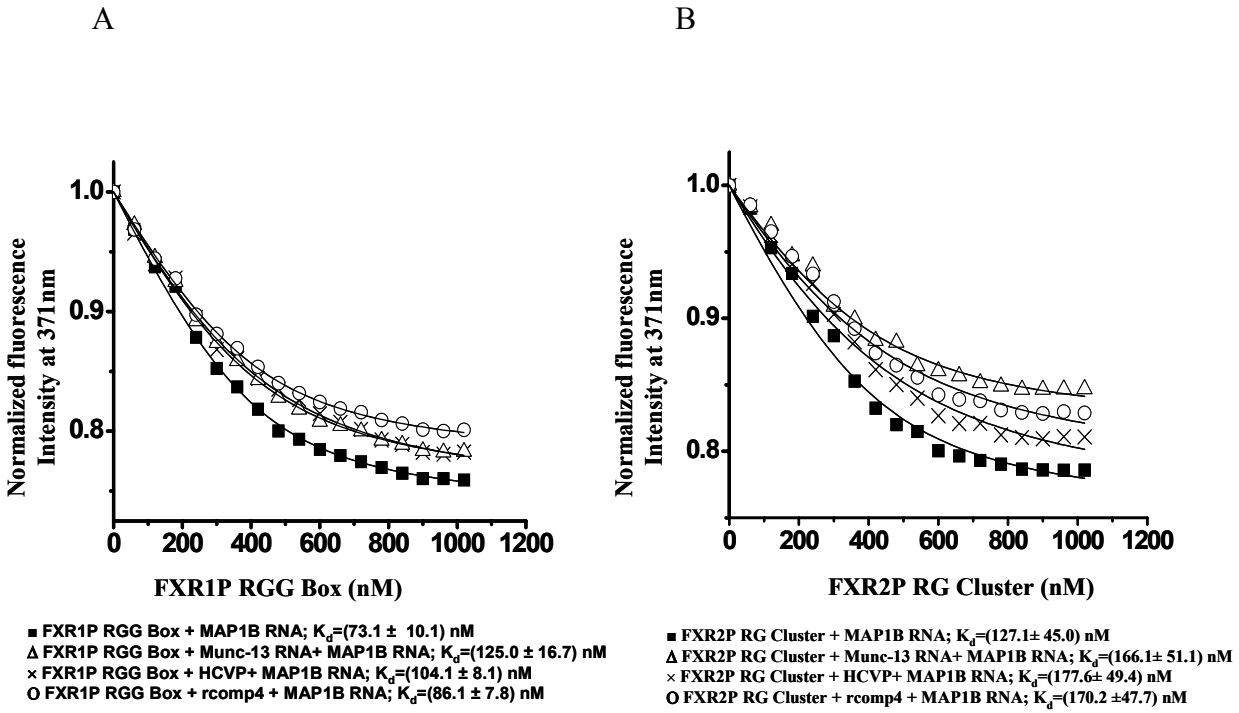

Figure 4.20 (A) Binding of the FXR1P RGG box to MAP1B-19AP RNA in the absence and presence of a 10-fold excess of Munc-13 site 1 RNA, rcomp4 single-stranded RNA , or HCV peptide **(B)** Binding of the FXR2P RG cluster to MAP1B-19AP in the absence and presence of a 10-fold excess of Munc-13 site 1 RNA , rcomp4 single-stranded RNA , or HCV peptide.

We wanted to analyze the effect of FXR1P RGG box and FXR2 RG cluster upon the G quadruplex structure of MAP1B RNA. Thus, we added each of the peptides to the RNA in a 1:1 ratio and monitored the effect of the peptides at 295 nm by UV spectroscopy. Like the FMRP RGG box, both the FXR1P RGG box and the FXR2 RG cluster were also found to slightly stabilize the MAP1B RNA G quadruplex structure (table 4.4).

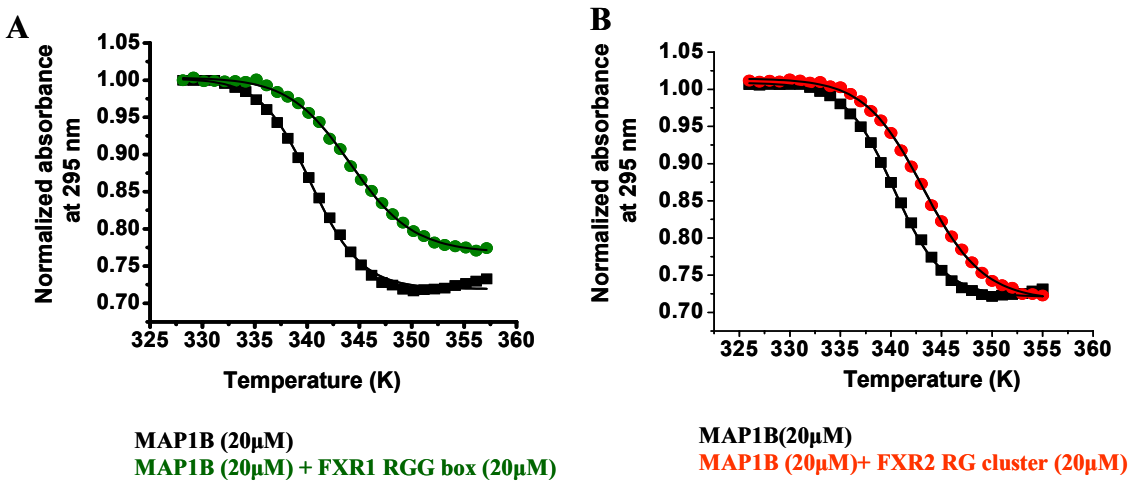

Figure 4.21 (A) UV melting profile of free 20 µM MAP1B RNA (black trace) and in a 1:1 ratio with the FXR1P RGG box (green trace). **(B)** UV melting profile of free 20 µM MAP1B RNA (blue trace) and in a 1:1 ratio with the FXR2P RG (red trace).

Molecule	Tm(^{0}C)	$\Delta H^0 f$ (kcal/mol)	$\Delta S^0 f$ (calmol^{-1}K^{-1})	$\Delta G^0 f$ (kcal/mol)
MAP1B	67.4 ± 0.1	-96.9 ± 0.1	-284.6 ± 0.2	- 12.1 ± 0.1
MAP1B+FXR1 RGG	70.8 ± 0.1	-83.7 ± 0.1	-243.4 ± 0.2	- 11.2 ± 0.1
MAP1B+FXR2 RG	70.0 ± 0.1	-91.2 ± 0.1	-265.8 ± 0.2	- 12.0 ± 0.1

Table 4.4 Thermodynamic parameters for the binding of FXR1P RGG box domain and the FXR2 RG cluster to the MAP1B RNA.

We have observed that when present in excess, the FMRP RGG box unwinds the MAP1B G quadruplex structure. To determine if the MAP1B RNA is affected in the same manner by the binding of FXR1P RGG box and of the FXR2P RG cluster, we titrated increasing amounts of each peptide to a fixed concentration of the RNA, monitoring the CD spectral changes. The G quadruplex structure of MAP1B RNA is completely unfolded by the presence of a 1:4 ratio of RNA: FXR1P RGG box, whereas it

remains unchanged even at a ratio of 1:10 RNA: FXR2 RG cluster (figure 4.22A and B). To ensure that the spectrum observed at 1:10 ratio of RNA: FXR1 RGG box is not due to RNA degradation, we performed a control experiment treating the sample containing a 1:10 ratio of RNA: FXR1 RGG box with proteinase K. We observed that upon the degradation of the RGG peptide, the G quadruplex structure of MAP1B RNA reforms, indicating that the spectral changes observed initially are not due to the RNA degradation (figure 4.23). It is noteworthy that the FXR1P RGG box has the ability to completely unfold the MAP1B RNA structure at only 1:4 RNA: peptide ratio as compared to the FMRP RGG box unwinding the MAP1B RNA at 1:8 ratio (32).

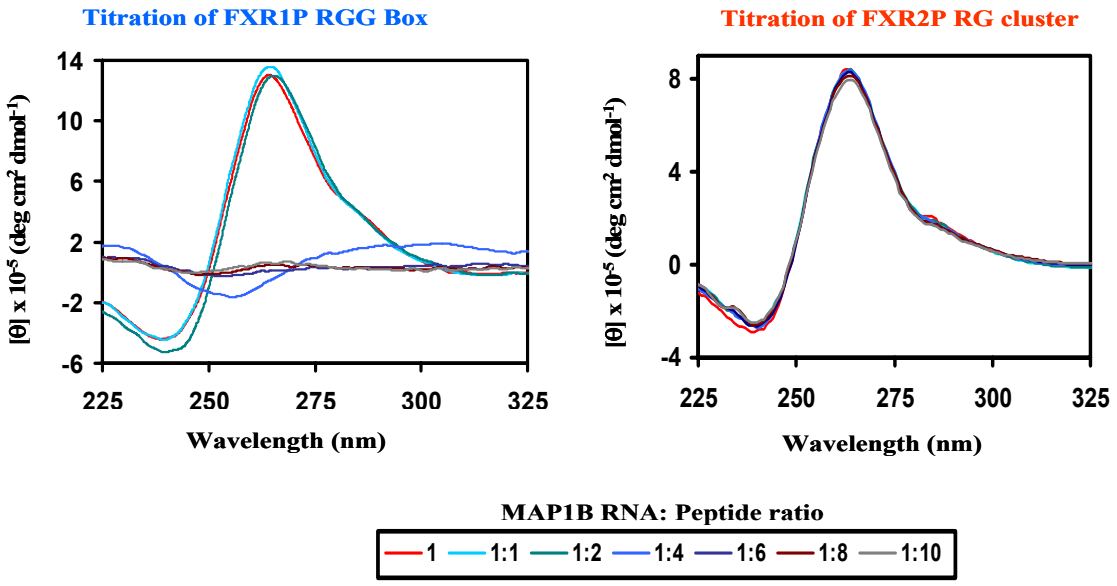

Figure 4.22 (A) CD spectra of MAP1B RNA in the presence of increasing concentrations of the FXR1P RGG box and **(B)** of the FXR2P RG cluster.

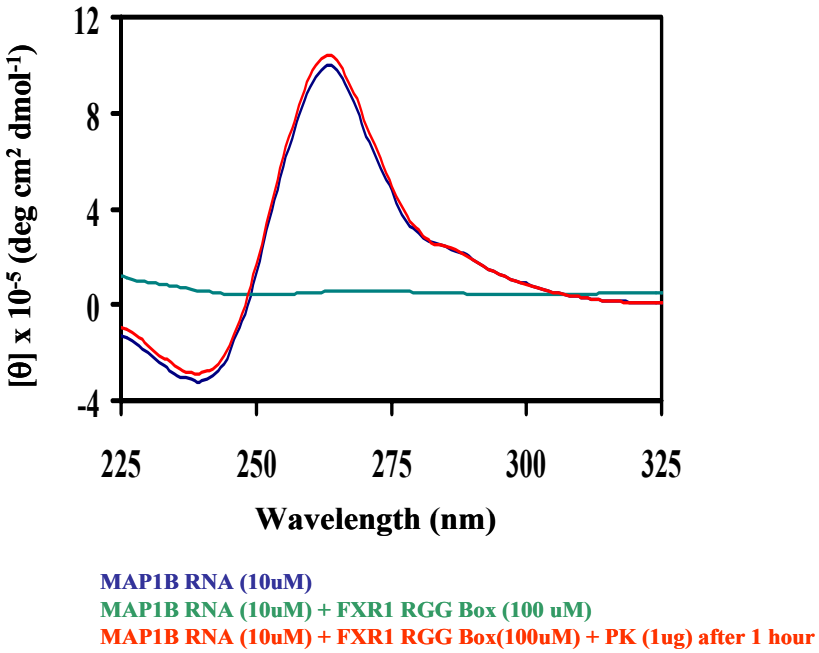

Figure 4.23 10 mM MAP1B RNA+ 100 mM FXR1P RGG Box with proteinase K (1 mg) for 1 h at 25°C, to check for the degradation of the RNA.

Chapter 5

5. Conclusions

RNA-binding proteins perform a variety of essential functions in the cell and regulate all aspects of RNA synthesis, such as mRNAs transcription, splicing, editing, and their translation into proteins. A striking example of the importance of RNA-binding proteins is the Fragile X Mental Retardation Protein (FMRP) - the loss of this protein, leads to the onset of Fragile X syndrome (FXS). The resulting effects are powerful, ranging from learning difficulty and hyperactivity to severe mental retardation in individuals. The cellular role of FMRP and how the absence of this protein causes mental retardation is poorly understood to this date (10). FMRP associates with polyribosomes and is involved in the translational repression of specific messenger RNA targets, important for neuronal development. The loss of this protein would cause defects in the proper translation of many mRNAs, thus explaining the pleiotropic phenotype associated with FXS. The specific mechanism by which FMRP recognizes its target RNAs and regulates their translation remains elusive, and has been a subject of continuous interest. Biochemical studies determined that FMRP binds with high affinity, using its RGG box RNA binding domain, to RNA sequences that have the potential to form G quadruplex

structures (44). This study contributes *to the understanding of the molecular basis of FMRP recognition of its mRNA targets and the role played by the G quadruplex structure in this process*. This is of particular interest in the context of fragile X syndrome research, because it contributes to our ability to identify other *in vivo* targets of FMRP.

The human semaphorin 3F mRNA and the microtubule associated protein 1B RNA have been identified both *in vivo* and *in vitro* as potential mRNA targets of FMRP and it has been proposed that their interactions with FMRP RGG box occur in a G quadruplex dependent manner (44).

We performed a detailed analysis of the FMRP interactions with S3F and MAP1B RNA targets by using molecular biology techniques and biophysical methods such as UV, circular dichroism, fluorescence and NMR spectroscopy. We demonstrated that both S3F and the MAP1B RNA adopt *in vitro intramolecular-parallel G quadruplex* structures capped by a stem structure. Our thermodynamic analysis is consistent with the presence of four G quartet planes in the G quadruplex structure of MAP1B RNA, coordinated by approximately four K^+ ion equivalents and of two G quartet planes in the S3F RNA structure (32, 47).

The properties of the RGG box RNA binding motif were then characterized in the context of *specific* binding to RNA. Dissecting the interactions between the FMRP RGG box and its mRNA targets revealed that this RNA binding motif binds specifically and with high affinity to the G quadruplex forming S3F and MAP1B RNA targets. At the physiological temperature the complex formation is enthalpically driven with an unfavorable entropic contribution. The analysis of the intricacy of the interactions, formed between the mRNA targets and the FMRP RGG box revealed that hydrophobic

effects are significant, whereas electrostatic interactions have only a minor contribution, to the overall stability of these complexes. *To our knowledge, these two studies are the first reports showing that the RGG box RNA binding domain binds specifically to an RNA target.*

Next, we investigated, if FMRP RGG box binding has any influence upon the G quadruplex structures of S3F and MAP1B RNA, our results indicating that in a 1:1 ratio the FMRP RGG box stabilizes the G quadruplex structures of both RNAs, to a larger extent for S3F than for MAP1B RNA. However, the addition of super stoichiometric amounts the FMRP RGG box leads to the unwinding of the G quadruplex structures of the S3F and MAP1B RNA and of another proposed mRNA target, namely Sc-1 RNA (data not shown). Based on these findings we propose a possible mechanism by which the control of FMRP concentrations in the cell might facilitate its translational regulator function.

FMRP has been proposed to be involved in the transport of specific mRNA targets to distant sites (as part of a large mRNP complex) where these mRNAs have to be maintained in a repressed state until synaptic input triggers their translation. According to our results, if the FMRP stabilization of the mRNA G quadruplex structure is associated with translation inhibition, then its unwinding could allow mRNA translation. Thus, a variation in the cellular concentration of FMRP, for example in response to a neurotransmitter stimulation event, could act as a regulatory switch for FMRP function, from translation repressor at low protein: RNA ratios, to translation promoter at high protein: RNA ratios (figure 5.1).

This model is supported by several studies showing that in response to application of glutamate or metabotropic glutamate receptor (mGluR) agonist (103,104), there was a rapid increase in FMRP synthesis locally in synaptoneurosomes indicating a regulation of FMRP concentration in response to synaptic stimulation. This local protein synthesis was then followed by a rapid decrease in FMRP levels back to baseline, suggesting a dynamic regulation of the FMRP concentration (104). Our model is thus consistent with this proposal where FMRP could bind and stabilize the G quadruplex structure present in a sub-set of its specific mRNA targets, maintaining them in a translationally repressed state (figure 5.1A). An increase in the FMRP concentration triggered by synaptic input, would lead to the unwinding of the G quadruplex structure of these mRNA targets, promoting their translation (figure 5.1B).

According to our model, the synaptically-triggered local protein synthesis from a G quadruplex containing class of FMRP mRNA targets would depend on the presence of a high FMRP concentration. Thus, the model predicts that in the absence of FMRP, the levels of these proteins will be unaffected by a neurotransmitter stimulation event. Consistent with this prediction, several studies have showed that the *Fmr1* knockout mice were deficient in their ability to induce local-protein synthesis of FMRP mRNA targets after stimulation of glutamate receptors (104-106) or of the NMDA-receptor (106).

A model which proposes the existence of a regulatory switch for the FMRP function from translation repressor to activator in response to the activation of mGluR has also been reported recently (33). These authors propose that the rapid degradation of FMRP observed after the stimulation of the mGluR, leads to the translation of the FMRP mRNA targets. This model (33) does not account for the observed initial surge in the

FMRP concentration (prior to its degradation and return to basal levels) in response to the stimulation of the mGluR. Our results suggests a functional role for the observed increase, followed by decrease in the FMRP concentration in response to neurotransmitter receptor stimulation, and the model we propose explains how such a switch of FMRP function could operate, at least for a subset of FMRP RNA targets that contain a G quadruplex structure.

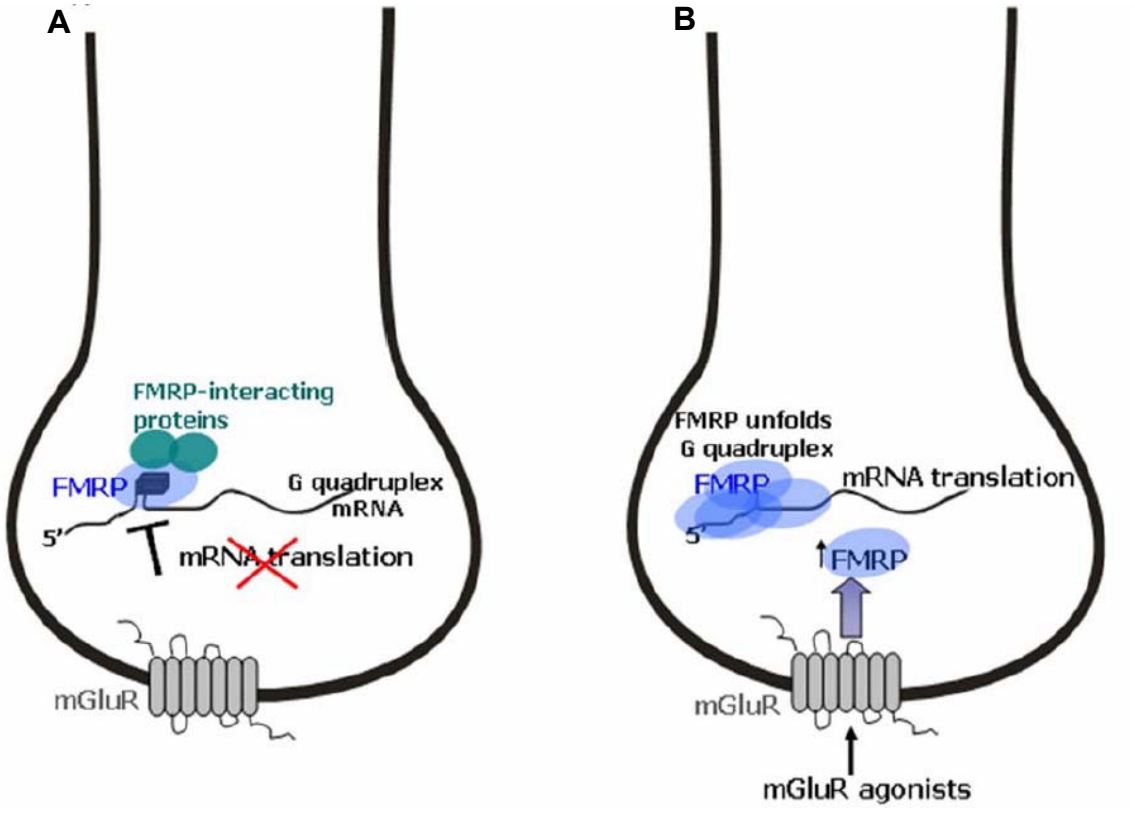

Figure 5.1 Proposed model for a regulatory switch of FMRP function from translation repressor to activator, in response to a neurotransmitter stimulation event. **(A)** In resting state when the FMRP:RNA ratio is low, FMRP represses the translation of specific mRNA targets by stabilizing a G quadruplex structure present in their 5'-UTR. **(B)** A stimulation of the mGluR by agonists leads to an increase in the FMRP concentration, unfolding the RNA G quadruplex structure, and allowing translation to occur.

The future goal of this project is to determine high-resolution structure of MAP1B RNA. We have already expressed and purified sufficient amounts of MAP1B RNA for NMR spectroscopy studies and we have established the experimental conditions in which this RNA forms a single species. We have confirmed that MAP1B RNA forms a G quadruplex structure by titrating increasing concentrations of KCl into a MAP1B RNA solution. High-resolution information about the interactions of MAP1B RNA and the FMRP RGG Box will be obtained using chemical shift perturbation methods.

The FMRP autosomal paralog, the FXR1P also specifically recognizes the S3F and the MAP1B RNA, albeit with a lower affinity as compared with FMRP. The RGG box of FXR1P acts in a similar fashion to the FMRP RGG Box by completely unwinding the G quadruplex structure of S3F and the MAP1B RNA. The FXR2 RG cluster recognizes specifically the MAP1B RNA but does not bind to the S3F mRNA. This RG cluster has no effect upon the G quadruplex structure of S3F and MAP1B RNA.

REFERENCES

1. Martin, J.B., and Bell, J. (1943) A pedigree of mental defect showing sex-linkage. *J. Neurol.Psychiatr.* **6**:154-7.

2. Lubs, H.A. (1969) A marker X chromosome. *Am J Hum Genet*.**21**:231-44.

3. Sutherland, G.R. (1977) Fragile sites on human chromosomes: demonstration of their dependence on the type of tissue culture medium. *Science.* **197**:265-6.

4. O'Donnell, W.T., and Warren, S.T. (2002) A decade of molecular studies of fragile x syndrome. *Annu. Rev. Neurosci*. **25**:315-38.

5. Penagarikano, O., Mulle, J. G., and Warren, S. T. (2007) The pathophysiology of fragile x syndrome, *Annu. Rev. Geno.human genetics 8*, 109-129.

6. Mohammad, G.(2000)Autism in mental retardation. *Curr. Opin. Psyc.* **13**:481-484.

7. Inoue, S.B., Siomi, M.C., and Siomi, H. (2000) Molecular mechanisms of fragile x syndrome. *J.Med. Invest*. **47**:101 -107.

8. Hagerman, R.J., Leehey, M., Heinrichs, W., Tassone, F., and Wilson, R. (2001) Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology*.**57**: 127-30.

9. Crawford, D.C., Acuna, J.M., and Sherman, S.L. (2001) FMR1 and the fragile x syndrome: human genome epidemiology review. *Genet. Med*. **3**: 359-371.

10. Jin, P., and Warren, S.T. (2003) New insights into fragile X syndrome: from molecules to neurobehaviors. *Tre. Bioch. Sci* .**28**: 152-158.

11. Jin, P., and Warren, S.T. (2000) Understanding the molecular basis of fragile X syndrome. *Hum. Mol. Gen*. **6**: 901-908.

12. Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F.P., Eussen, B.E., van Ommen G.J.B., Blonden, L.A.J., Riggins, G.J., Chastain, J.L., Kunst, C.B., Galijaard, H., Caskey, C.T., Nelson, D.L., Oostra, B.A and Warren, S.T. (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell.* **65:**905–914.

13. Imbert, G., Feng, Y., Nelson, D., Warren, S.T and Mandel, J.L. (1998) FMR1 and mutations in fragile X syndrome: molecular biology, biochemistry and genetics.

14. Warren, S.T. and Wells, R.D. (Eds) Genetic instabilities and hereditary neurological diseases. *Academic press*: 27-53.

15. Pieretti, M., Zhang, F., Fu, Y.H., Warren, S.T., Oostra, B.A., Caskey, C.T and Nelson, D.L. (1991) Absence of expression of FMR1 gene in fragile x syndrome. *Cell.* **66**: 817-822.

16. Mandel, J.L. and Heitz, D. (1992) Molecular genetics of the fragile x syndrome: A novel type of unstable mutation. *Curr Opin Gen Dev.* **2**:422-430.

17. Hagerman, R.J., Ono, M.Y., and Hagerman, P.J. (2005) Recent advances in fragile X: a model for autism and neurodegeneration. *Curr OpinPsyc.* **18(5):** 490-6.

18. Cummings, C.J., and Zoghbi, H.Y. (2000) Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum.Mol.Gen.* 9(6): 909-16.

19. Eichler EE, Richard S, Gibbs RA and Nelson DL (1993) Fine structure of human FMR 1gene. Hum.Mol.Genet.2:1147-53.

20. Ashley CT, Wilkinson KD, Reines D and Warren ST. (1993) FMR1 protein: conserved RNP family domains and selective RNA binding. *Scienc*e.**26**2: 563-66.

21. Wan L, Dockendorff TC, Jongens TA, Dreyfuss G. (2000) Characterization of dFMR1, a drosophila melanogaster homolog of the fragile X mental retardation protein. *Mol.Cell. Biol*.**20**:8536-8547.

22. Devys D, Lutz Y, Rouyer N, Bellocq JP and Mandel JL. (1993) FMR-1 protein is cytoplasmic, is most abundant in neurons and appears normal in carriers of fragile X premutation. *Nat. Gen*. **4**, 335-340.

23. Khandjian, E.W. (1999) Biology of the fragile x mental retardation protein, an RNA – binding protein. *Biochem. Cell. Biol* .**77**:331-342.

24. Siomi H, Siomi MC, Nussbaum RL and Dreyfuss G. (1994) The protein product of fragile x gene , FMR 1 has characteristics of an RNA binding protein. *Cell.* **74:**291-98.

25. Mazroui R, Huot ME, Tremblay S, Boilard N, Labelle Y, and Khandjian EW. (2003) Fragile X Mental Retardation protein determinants required for its association with polyribosomal mRNPs. *Hum. Mol. Genet.* **12**: 3087-3096.

26. Adinolfi S, Bagni C, Musco G, Gibson T, Mazzarella L and Pastore A. (1999) Dissecting FMR1, the protein responsible for fragile X syndrome, in its structural and functional domains. *RNA*. **9**:1248-58.

27. Antar LN and Bassell GJ. (2003) Sunrise at the synapse: the FMRP mRNP shaping the synaptic interface. *Neuron.* **37**: 555-558.

28. Massimiliano V, Zalfa F and Bagni C. (2004) FMRP and its target RNAs: fishing for the specificity. *Neuroreport*. **16**:2447-2450.

29. Laggerbauer B, Ostareck D, Keidel EM, Lederer AO and Fischer U. (2001) Evidence that fragile X mental retardation protein is a negative regulator of translation. *Hum.Mol. Gen*. **4**: 329 -338.

30. Mazroui R, Huot ME, Tremblay S, Boilard N, Labelle Y and Khandjian EW. (2002) Trapping of messenger RNA by fragile X mental retardation protein inhibits translation via interacting with mRNA. *Hum. Mol. Genet.* **11**: 3007-3017.

31. Li Z, Zhang Y, Ku L, Wilkinson KD, Warren ST and Feng Y. (2001) The fragile X mental retardation protein inhibits translation via interacting with an mRNA. *Nucleic Acids Res*. **29**: 2276-2283.

32. Menon, L., Mader, S.A., and Mihailescu, M.R.(2008) Fragile X Mental Retardation Protein interactions with the Microtubule associated protein 1B RNA, *RNA*, In press.

33. Ronesi, J.A., and Huber, K.M. (2008) Metabotropic glutamate receptors and fragile X mental retardation protein: partners in translational regulation at the synapse. *Science Signaling* **1(5)**, pe6.

34. Gabus C, Mazroui R, Tremblay S, Khandjian EW and Darlix JL. (2004) The fragile X mental retardation protein has nucleic acid chaperone properties. *Nucleic Acids Res.* **7**, 2129-2137.

35. Siomi H, Siomi MC, Nussbaum RL and Dreyfuss G. (1993). The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein. *Cell.* **2**:291-8.

36. Bagni, C. and Greenough W.T.(2005) From mRNP trafficking to spine dysmorphogenesis: the roots of fragile X syndrome. *Nat. Rev. Neurosci.* **6**, 376-387.

37. Siomi H, Matunis MJ, Michael WM and Dreyfuss G. (1993) The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res*. **5**:1193-8.

38. Brown V, Small K, Lakkis L, Feng Y, Gunter C, Wilkinson KD, and Warren ST. (1998) Purified Recombinant Fmrp Exhibits Selective RNA Binding as an Intrinsic Property of the Fragile X Mental Retardation Protein. *J Biol Chem*. 273:15521-15527.

39. Kiledjian, M., and Dreyfuss, G. (1992) Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J*. 11(7): 2655-64.

40. Zalfa F, Giorgi M, Primerano B, Moro A, Di Penta A, Reis S, Oostra B, and Bagni C. (2003) The Fragile X Syndrome Protein FMRP Associates with BC1 RNA and Regulates the Translation of Specific mRNAs at Synapses. *Cell*. **112**: 317-327.

41. Iacoangeli A, Rozdestvensky TS, Dolzhanskaya N, Tournier B, Schutt J, Brosius J, Denman RB, Khandjian EW, Kindler S, Tiedge H (2008). On BC1 RNA and the fragile X mental retardation protein. *Proc Natl Aca Sci*, **15**; 105(2): 734-9.

42. Iacoangeli A, Rozdestvensky TS, Dolzhanskaya N, Tournier B, Schutt J, Brosius J, Denman RB, Khandjian EW, Kindler S, Tiedge H (2008). Reply to Bagni:On BC1 RNA and the fragile X mental retardation protein. *Proc Natl Aca Sci*, **15**; 105(2): 734-9.

43. Schaeffer C, Bardoni B, Mandel JL, Ehresmann B, Ehresmann C. and Moine H.(2001) The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J.* **20**: 4803-4813.

44. Darnell JC, Jensen KB, Jin P, Brown V, Warren S.T, and Darnell RB. (2001) Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell*. **107:** 489-499.

45. Ramos A, Hollingworth D and Pastore A. (2003) G-quartet-dependent recognition between the FMRP RGG box and RNA. *RNA*. **9:** 1198.

46. Zanotti, K.J., Lackey, P.E., Evans, G.L., and Mihailescu, M.R. (2006) Thermodynamics of the fragile X mental retardation protein RGG box interactions with G quartet forming RNA. *Biochemistry* **45** (27), 8319 -8330.

47. Menon, L., and Mihailescu, M.R. (2007) Interactions of the G quartet forming semaphorin 3F RNA with the RGG box domain of the fragile X protein family. *Nucleic Acids Res*. **35**(16), 5379-92.

48. Ceman S, Brown V, and Warren ST. (1999) Isolation of an FMRP-associated messenger ribonucleoprotein particle and identification of nucleolin and the fragile X-related proteins as components of the complex. *Mol. Cell Biol*. **19**:7925-7932.

49. Brown V, Jin P, Ceman S, Darnell JC, O'Donnell WT, Tenenbaum SA, Jin X, Feng Y, Wilkinson KD, Keene JD, Darnell RB, and Warren ST. (2001) Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell*. **107**:477-487.

50. Miyashiro KY, Mitchener BA., Purk TP, Becker KG, Barret T, Liu L, Carbonetto S, Weiler IJ, Greenough WT and Eberwine J. (2003) RNA cargoes associating with FMRP reveal deficits in cellular functioning in Fmr1 null mice. *Neuron.* **37**: 417-431.

51. Shafer RH and Smirnov I. (2001) Biological aspects of DNA/RNA quadruplexes. *Biopoly.* **56**: 209-227.

52. Davis JT. (2004) G-quartets 40 years later: from 5'GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed*, **43**: 668-698.

53. Awang G and Sen D (1993) Mode of dimerization of HIV-1 genomic RNA. *Biochem* 32 (42):11453-7.

54. Bonnal S, Schaeffer C, Creancier L, Clamens S, Moine H, Prats AC, Vagner S (2003) A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *J Biol Chem*. 278(41):39330-6.

55. Gomez D, Lemarteleur T, Lacroix L, Mailliet P, Mergny JL and Riou JF (2004) Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Res*. 32(1): 371-9.

56. Kostadinov R, Malhotra N, Viotti M, Shine R, D'Antonio and Bagga P (2006) GRSDB: a database of quadruplex forming G-rich sequences in alternatively processed mammalian pre-mRNA sequences. *Nucleic Acids Res*. 34: D119-24.

57. Kumari S, Bugaut A, Huppert JL and Balasubramaniam S (2007) An RNA G-quadruplex in the 5'UTR of the NRAS proto-oncogene modulates translation. *Nat Chem Biol* **3(4)**:218-21.

58. Patel DJ, Phan AT and Kuryavyi V (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res*. 35(22): 7429-55.

59. Williamson, J.R. (1994) G-quartet structures in telomeric DNA. *Annu. Rev. Biophys. Biomol. Struc.* **23**,703-30.

60. Gellert M, Lipsett MN and Davies DR. (1962) Helix formation by guanylic acid. *Proc.Natl.Acad.Sci. USA* **48**: 2013-18.

61. Hazel, P., Huppert, J., Balasubramaniam, S., and Neidle, S. (2004) Loop- length-dependent folding of G-Quadruplexes. *J. Am. Chem. Soc.*, **126**, 16405-16415.

62. Han HY and Hurley LH. (2000) G-quadruplex DNA: a potential target for anti-cancer drug design. *Trends Pharmacol. Sci.* **21**:136–142.

63. Mergny JL, Cian AN, Ghelab A, Sacca B and Lacroix L. (2005) Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res*. **33**: 81-94.

64. Nasarre, P., Constantin, B., Rouhaud, L., Harnois, T., Raymond, G., Drabkin, H. A., Bourmeyster, N., and Roche, J. (2003) Semaphorin SEMA3F and VEGF have opposing effects on cell attachment and spreading, *Neoplasia (New York, N.Y 5*, 83-92.

65. Guan, F., Villegas, G., Teichman, J., Mundel, P., and Tufro, A. (2006) Autocrine class 3 semaphorin system regulates slit diaphragm proteins and podocyte survival, *Kidney international 69*, 1564-1569.

66. Xiang R, Davalos AR, Hensel CH, Zhou XJ, Tse C and Naylor SL. (2002) Semaphorin 3F Gene from Human 3p21.3 Suppresses Tumor Formation in Nude Mice. *Can. Res.* **62**, 2637-2643.

67. Kolodkin AL, Matthes DJ, O'Connor TP, Patel NH, Admon A, Bentley D and Goodman C S. (1992) Fasciclin IV: sequence, expression, and function during growth cone guidance in the grasshopper embryo. *Neuron*. **9***:* 831-845.

68. Rackham O and Chris B. (2004) Visualization of RNA-protein interactions in living cells: FMRP and IMP1 interact on mRNAs. *EMBO J*. **23**: 3346-3355.

69. Gonzalez-Billault, C., Jimenez-Mateos, E.M., Caceres, A., Diaz-Nido, J., Wandosell, F., and Avila, J. (2004) Microtubule-associated protein 1B function during normal development, regeneration, and pathological conditions in the nervous system. *Neurobiol*. **58**(1),48-59.

70. Zhang, Y.Q., Bailey, A.M., Matthies, H.J., Renden, R.B., Smith, M.A., Speese, S.D., Rubin, G.M.,and Broadie, K.(2001) Drosophila fragile X-related gene regulates the MAP1B homolog Futsch to control synaptic structure and function. *Cell* **107**(5):591-603.

71. Lu, R., Wang, H., Liang, Z., Ku, L., O'Donnell, W.T., Li, W., Warren, S.T., and Feng, Y. (2004) The fragile X protein controls microtubule-associated protein 1B translation and microtubule stability in brain neuron development. *Proc. Natl. Acad. Sci. U. S. A.* **101**(42), 15201 – 15206.

72. Antar, L.N., Dictenberg, J.B., Plociniak, M., Afroz, R., and Bassell, G.J. (2005) Localization of FMRP-associated mRNA granules and requirement of microtubules for activity-dependent trafficking in hippocampal neurons. *Genes Brain Behav.* **4**, 350-359.

73. Zhang, Y., O'Connor, J.P., Siomi, M.C., Srinivasan, S., Dutra, A., Nussbaum, R.L. and Dreyfuss, G. (1995) The fragile X mental retardation syndrome protein interacts with novel homologs FXR1and FXR2. *EMBO J*., **14**, 5358–5366.

74. Tamanini, F., Van Unen, L., Bakker, C., Sacchi, N., Galjaard, H.,Oostra, B.A. and Hoogeveen, A.T. (1999) Oligomerization properties of fragile-X mental-retardation protein (FMRP) and the fragile-X-related proteins FXR1P and FXR2P. *Biochem. J*., **343**, 517–523.

75. Siomi, M.C., Zhang, Y., Siomi, H. and Dreyfuss, G. (1996) Specific sequences in the fragile X syndrome protein FMR1 and the FXR proteins mediate their binding to 60S ribosomal subunits and the interactions among them. *Mol. Cell. Biol,* **16**, 3825-3832.

76. Tamanini, F., Willemsen, R., van Unen, L., Bontekoe, C., Galjaard, H., Oostra, B. A., Hoogeveen, A. T. (1997) Differential expression of FMR1, FXR1 and FXR2 proteins in human brain and testis. *Hum. Molec. Genet,* **6**, 1315-1322, 1997.

77. Bardoni B, Schenck A, and Mandel JL.(1999) A novel RNA-binding nuclear protein that interacts with the fragile X mental retardation protein. *Hum.Mol.Gen* 2557-2566.

78. Milligan, J.F. & Uhlenbeck, O.C. (1989), Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol*, **180**, 51-62.

79. Mergny, J.-L., Phan, A.-T. & Lacroix, L., (1998) Following G-quartet formation by UV- spectroscopy. *FEBS Lett*. **435**, 74-78.

80. Becktel, W. J., and Schellman, J. A. (1987) Protein Stability Curves. *Biopolymers* **26,** 1859-1877.

81. Ha, J.-H., Spolar, R.S., and Record, M.T. (1989) Role of the hydrophobic effect in stability of site-specific protein-DNA complexes. *J. Mol. Biol.* **209**, 801-816.

82. Plateau, P. & Gueron, M., (1982) Exchangeable protons without base line distortion using a new strong pulse sequence. *J. Am. Chem. Soc*., **104**, 7310-7311.

83. Piotto, M., Saudek, V., and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR.* **2**: 661–665.

84. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406-3415.

85. Dolinnaya, N. G. and Fresco, J. R.(1992) Single-stranded nucleic acid helical secondary structure stabilized by ionic bonds: d(A(+)-G)10. *Proc Natl Acad Sci* U S A. **89**(19): 9242–9246.

86. Shiber, M.C., Braswell, E.H., Klump, H., Fresco, J.R.(1996) Duplex-tetraplex equilibrium between a hairpin and two interacting hairpins of d(A-G)10 at neutral pH. *Nucleic Acids Res.,* **24**(24):5004–5012.

87. Hardin, C. G., Perry, A. G., and White, K. (2001) Thermodynamic and kinetic characterization of the dissociation and assembly of quadruplex nucleic acids, *Biopolymers* 56, 147-194.

88. Gaffney, B.L., Wang, C., and Jones, R.A. (1992) Tetraplex formation of d[G*(* ^{15}N 7) G T T T T T G G] and d[T(^{15} N 7) G G G T] monitored by proton detected ^{15} N NMR. *J. Am. Chem. Soc*. **114**, 4047-4050.

89. Jin, R., Gaffney, B.L., Wang, C., Jones, R.A., Breslauer and Kenneth, J. (1992) Thermodynamics and Structure of a DNA Tetraplex: A Spectroscopic and Calorimetric Study of the Tetramolecular Complexes of d(TG_3T) and d($TG_3T_2G_3T$). *Proc. Natl. Acad. Sci. U. S. A.* **89**, 8832-36.

90. Seenisamy, J., Rezler, E.M., Powell, T.J., Tye, D., Gokhale, V., Joshi, C.S., Siddiqui-Jain, A., and Hurley, L.H. (2004) The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J. Am. Chem. Soc.* **126**, 8702–8709.

91. Ambrus, A., Chen, D., Dai, J., Jones, R.A., and Yang, D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization*., Biochemistry* **44**, 2048–2058.

92. Matsugami**,**A., Ouhashi,K., Kanagawa,M., Liu,H., Kanagawa,S., Uesugi,S., and Katahira, M. (2001) **A**n intramolecular quadruplex of (GG**A**)(4) triplet repeat DN**A** with a G:G:G:G tetrad and a G(:**A**):G(:**A**):G(:**A**):G hept**a**d, and its dimeric interaction. *J. Mol. Biol.* **313**, 255–269.

93. Matsugami, A., Okuizumi, T., Uesugi, S., and Katahira, M. (2003). Intramolecular higher order packing of parallel quadruplexes comprising a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad of GGA triplet repeat DNA. *J. Biol. Chem.* **278**(30):28147-53.

94. Xu, Y., Noguchi, Y., and Sugiyama, H. (2006) The new models of the human telomere d [AGGG ($TTAGGG)_3$] in K^+ solution . *Bioorg. Med. Chem.* **14**, 5584–5591.

95. Pedroso, I.M., Duarte, L.F., Yanez, G., Burkewitz, K., and Fletcher, T.M. (2007) Sequence specificity of inter- and intramolecular G-quadruplex formation by human telomeric DNA. Biopolymers. *Biopolymers* **87**(1), 74-84.

96. Cantor, C. R., and Schimmel, P. R. (1980) *Biophysical Chemistry*, Parts I-III, NewYork, W. H. Freeman and Co.

97. Spolar, R.S., and Record, M.T. (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263**, 777-784.

98. Record, M.T., Ha, J.H., and Fisher, M. (1991) Use of equilibrium and kinetic measurements to determine the thermodynamic origins of stability and specificity and mechanism of formation of site specific complexes between proteins and helical DNA. *Methods Enzymol.* **208**, 291-343.

99. Draper, D.E. (1995) Protein-RNA recognition. *Annu. Rev. Biochem*. 64, 593-620.

100. Baumann, C., Otridge, J., and Gollnick, P. (1996) Kinetic and Thermodynamic Analysis of the Interaction between TRAP (trp RNA-binding Attenuation Protein) of Bacillus subtilis and trp Leader RNA. *J. Biol. Chem.* **271**,12269–12274.

101. Fisher, B.M., Ha, J-H., and Raines, R.T. (1998) Coulombic forces in protein-RNA interactions: Binding and cleavage by ribonuclease A and variants at Lys7, Arg10, and Lys66. *Biochemistry* **37,** 12121–12132.

102. Wilson, G.M., Sutphen, K., Bolikal, S., Chuang, K.Y., and Brewer, G (2001) Thermodynamics and kinetics of Hsp70 association with A + U-rich mRNA-destabilizing sequences. *J. Biol. Chem.* **276**(48),44450-44456.

103. Weiler, I.J., Irwin, S.A., Klintsova, A.Y., Spencer, C.M., Brazelton, A.D., Miyashiro, K., Comery, T.A., Patel, B., Eberwine, J., and Greenough, W.T. (1997) Fragile X mental retardation protein is translated near synapses in response to neurotransmitter activation. *Proc. Natl. Acad. Sci. U. S. A.* **94** , 5395-5400.

104. Hou, L., Antion, M.D., Hu, D., Spencer, C.M., Paylor, R., and Klann, E. (2006) Dynamic translational and proteasomal regulation of fragile X mental retardation protein controls mGluR-dependent long-term depression. *Neuron* **51**,441–454.

105. Todd, P.K., Mack, K.J., and Malter, J.S.(2003) The fragile X mental retardation protein is required for type-I metabotropic glutamate receptor-dependent translation of PSD-95.*Proc. Natl. Acad. Sci. U. S. A.* **100** , 14374–14378.

106. Muddashetty, R.S., Kelić, S., Gross, C., Xu, M., and Bassell, G.J. (2007) Dysregulated metabotropic glutamate receptor-dependent translation of AMPA Receptor and postsynaptic density-95 mRNAs at synapses in a mouse model of fragile X syndrome. *J. Neurosci.* **27**(20), 5338-5348.